25 to about amino acid 250 of the full length human erythropoietin receptor protein, said human erythropoietin receptor polypeptide being capable of binding human erythropoietin, wherein said polypeptide is non-glycosylated.

c2 cont.

REMARKS

Applicant confirms the election of group II, claim 3 and 5, made on May 13, 1998. Claims 3 and 5 are pending in the application.

Applicant has amended the specification to include a priority claim to the parent application, U.S. Serial No. 08/106,815, which was filed on August 16, 1993.

Applicant has amended independent claim 3 to recite that the polypeptide is unglycosylated, as the human erythropoietin receptor polypeptide of the invention was expressed in *E. coli*. One of ordinary skill in the art at the time the invention was made would have appreciated that a prokaryotically expressed polypeptide does not contain post-translational modifications such as glycosylation. See, for example, page 16.3 of Sambrook et al., Molecular Cloning: A Laboratory Manual, 1989; and Bassuk et al., Arch. Biochem. Biophys., 1996, 325(1):8-19, copies of which are attached hereto.

No new matter is added by these amendments.

Rejections under 35 U.S.C. §102 and §103

The Examiner rejected claims 3 and 5 under 35 U.S.C. §102(b) as being anticipated by D'Andrea et al., U.S. Patent No.

5,378,808 (the '808 patent). Claim 5 also was rejected under 35 U.S.C. §103 as being unpatentable over the '808 patent. The Examiner asserted that D'Andrea describes production and purification of the secreted human EPO receptor polypeptide. The Examiner also asserted that it would have been obvious to one of ordinary skill in the art to "employ the secreted receptor polypeptide described by D'Andrea to make an affinity chromatography matrix capable of binding EPO".

Amended claim 1 recites that the purified human erythropoietin receptor polypeptide is unglycosylated. The '808 patent does not disclose production of a purified, unglycosylated, human erythropoietin receptor polypeptide, as the polypeptide of the '808 patent was expressed in COS cells. One of ordinary skill in the art would appreciate that COS cells are eukaryotic cells that glycosylate expressed proteins. See, for example, page 8576 of the attached article by Aruffo and Seed, Proc. Natl. Acad. Sci. USA, 1987, 84:8573-8577.

As the '808 patent does not disclose production of the unglycosylated human erythropoietin receptor polypeptide and immunoassay composition of the present invention, claims 3 and 5 are not anticipated by the '808 patent. In addition, since the glycosylation state of a protein can have significant impact on protein interactions, the '808 patent does not render claim 5 obvious. See, for example, Roos et al., Mol. Cell. Biol., 1997, 17(11):6472-6480, a copy of which is attached hereto. In light of this, the Examiner is requested to withdraw the rejection of

claims 3 and 5 under 35 U.S.C. § 102 and § 103 over the '808 patent.

Applicant submits that all of the claims are now in condition for allowance, which action is requested. The Examiner is invited to telephone the undersigned if it is felt that such would advance prosecution of the application.

Please charge any additional fees, or make any credits, to Deposit Account No. 06-1050.

Respectfully submitted,

Date: March 22, 1999

Mark S. Ellinger, Ph.D.
Reg. No. 34,812

Fish & Richardson P.C., P.A.
60 South Sixth Street, Suite 3300
Minneapolis, MN 55402

Telephone: 612/335-5070
Facsimile: 612/288-9696
20945.M11

3

# Molecular Cloning

## A LABORATORY MANUAL

## SECOND EDITION

**J. Sambrook**
UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER

**E.F. Fritsch**
GENETICS INSTITUTE

**T. Maniatis**
HARVARD UNIVERSITY


*Cold Spring Harbor Laboratory Press*
*1989*

***Molecular Cloning***

A LABORATORY MANUAL

SECOND EDITION


Printed in the United States of America

9 8 7 6 5 4 3

*Book and cover design by Emily Harste*

*Cover:* The electron micrograph of bacteriophage λ particles stained with uranyl acetate was digitized and assigned false color by computer. (Thomas R. Broker, Louise T. Chow, and James I. Garrels)

*Cataloging in Publications data*

Sambrook, Joseph
Molecular cloning : a laboratory manual / E.F. Fritsch, T. Maniatis—2nd ed.
p. cm.
Bibliography: p.
Includes index.
ISBN 0-87969-309-6
1. Molecular cloning—Laboratory manuals. 2. Eukaryotic cells--Laboratory manuals. I. Fritsch, Edward F. II. Maniatis, Thomas III. Title.
QH442.2.M26 1987
574.87'3224—dc19 87-35464



All Cold Spring Harbor Laboratory Press publications may be ordered directly from Cold Spring Harbor Laboratory Press, 10 Skyline Drive, Plainview, New York 11803. Phone: 1-800-843-4388. In New York (516) 367-8423. FAX: (516) 367-8432.

# Expression of Proteins

## EXPRESSION OF PROTEINS FROM CLONED GENES

A few eukaryotic proteins have been expressed efficiently and inexpensively in prokaryotic hosts (see Chapter 17). However, many eukaryotic proteins synthesized in bacteria fold incorrectly or inefficiently and, consequently, exhibit low specific activities. In addition, production of authentic, biologically active eukaryotic proteins from cloned DNA frequently requires posttranslational modifications such as accurate disulfide bond formation, glycosylation, phosphorylation, oligomerization, or specific proteolytic cleavage—processes that are not performed by bacterial cells. This problem is particularly severe when expression of functional membrane or secretory proteins such as cell surface receptors and extracellular hormones or enzymes is required.

Because of these problems, considerable effort has been made to develop systems to express mammalian proteins in mammalian cells. These systems can be divided into two types: those that involve transient or stable expression of transfected DNA and those that involve the use of viral expression vectors derived from simian virus 40 (SV40) (Elder et al. 1981; Gething and Sambrook 1981; Rigby 1982, 1983; Doyle et al. 1985; Sambrook et al. 1986), vaccinia virus (Mackett et al. 1985; Moss 1985; Fuerst et al. 1986, 1987), adenovirus (Solnick 1981; Thummel et al. 1981, 1982, 1983; Mansour et al. 1985; Karlsson et al. 1986; Berkner 1988), retroviruses (Dick et al. 1986; Gilboa et al. 1986; Eglitis and Anderson 1988), and baculoviruses (Luckow and Summers 1988). The diversity of these animal viruses is so great that an account of their molecular biology is beyond the scope of this chapter. In addition, effective utilization of viral vectors requires some prior experience in the methods used to grow, quantitate, and plaque-purify different viruses. An entire chapter would be required to provide all of the information necessary to use each of these vectors. For these reasons, we have chosen to focus entirely on expression methods that involve DNA transfection. (*Note*: An excellent manual describing methods for the use of baculovirus vectors and procedures for culture of insect cells has been published by Summers and Smith [1987].)

Expression of proteins from cloned eukaryotic genes in mammalian cells has been used for a number of different purposes:

- To confirm the identity of a cloned gene by using immunological or functional assays to detect the encoded protein
- To express genes encoding proteins that require posttranslational modifications such as glycosylation or proteolytic processing
- To produce large amounts of proteins of biological interest that are normally available in only limited quantities from natural sources
- To study the biosynthesis and intracellular transport of proteins following their expression in various cell types

Natl. Acad. Sci. USA
84, pp. 8573–8577, December 1987
Immunology



# Molecular cloning of a CD28 cDNA by a high-efficiency COS cell expression system

(expression library/surface antigen/cloning strategy/Tp44)

ALEJANDRO ARUFFO AND BRIAN SEED

Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

*Communicated by Richard Axel, July 10, 1987*

**ABSTRACT** **CD28 (Tp44) is a human T-cell-specific homodimer surface protein that may participate in T-cell activation. We have isolated a cDNA clone encoding CD28 by a simple and highly efficient cloning strategy based on transient expression in COS cells. Central to this strategy is the use of an efficient method to prepare large plasmid cDNA libraries. The libraries are introduced into COS cells, where transient expression of surface antigen allows the isolation of cDNAs by way of monoclonal antibody binding. The CD28 cDNA encodes a highly glycosylated membrane protein with homology to the immunoglobulin superfamily and directs the production of a homodimer in transfected COS cells.**

Monoclonal antibodies recognizing three antigens, CD3 (T3), CD2 (T11), and CD28 (Tp44), cause human T cells to proliferate in the presence of phorbol esters (1–3). Whereas CD3 appears to be involved in transduction of the signal generated by antigen binding to the T-cell receptor (4, 5), the role of the CD2 and CD28 antigens in physiological proliferation is not at present understood. Exposure of T cells to anti-CD28 or interleukin 1 results in increased levels of cytoplasmic cGMP (6, 7), and either interleukin 1 or anti-CD28 antibody can substitute for monocytes in provoking proliferation of T cells exposed to agarose-bound anti-CD3 antibody (7, 8). However, the molecular weight, tissue distribution, and surface density of the interleukin 1 receptor (9) and CD28 antigen (10) are dissimilar. Moreover, anti-CD28 treatment, but not interleukin 1 treatment, allows proliferation of phorbol ester-stimulated T cells in the presence of dibutyryl cAMP (11).

We have described (12) a monoclonal antibody-based technique for enrichment of cDNAs encoding surface antigens. Here we describe a method of constructing plasmid expression libraries that allows the enrichment technique to be fully exploited. The method for making plasmid expression libraries may be of more general use for expression cloning, since, with the exception of some lymphokine cDNAs isolated by expression in COS cells (13–16), few cDNAs in general have been isolated from mammalian expression libraries. There appear to be two principal reasons for this. (*i*) The existing technology (17) for construction of large plasmid libraries is difficult to master, and the library size, even in the hands of virtuosos, rarely approaches that accessible by phage cloning techniques (18). (*ii*) The existing vectors are, with one exception (13), poorly adapted for high-level expression, particularly in COS cells. The reported successes with lymphokine cDNAs do not imply a general fitness of the methods used, since these cDNAs are particularly easy to isolate from expression libraries. Lymphokine bioassays are very sensitive (13–16), and the mRNAs are typically both abundant and short (13–16).

In this article we have applied the antibody selection technique to isolate a cDNA clone encoding the CD28 antigen.* The antigen shares substantial homology with members of the immunoglobulin superfamily and forms a dimer structure on the surface of transfected COS cells similar to the dimer structure found on T lymphocytes.

## MATERIALS AND METHODS

**Preparation of cDNA Libraries.** Poly(A)$^+$ RNA was prepared from the human T-cell tumor line HPB-ALL by oligo(dT)-cellulose chromatography of total RNA isolated by the guanidine thiocyanate method (19). cDNA was prepared by the following protocol, based on the method of Gubler and Hoffman (20). mRNA (4 $\mu$g) was heated to ≈100°C in a 1.5-ml centrifuge tube for 30 sec and quenched on ice, and the volume was adjusted to 70 $\mu$l with RNase-free water. To this were added 20 $\mu$l of buffer [0.25 M Tris, pH 8.8 (pH 8.2 at 42°C)/0.25 M KCl/30 mM $MgCl_2$], 2 $\mu$l of RNase inhibitor (Boehringer Mannheim; 36 units/$\mu$l), 1 $\mu$l of 1 M dithiothreitol, 1 $\mu$l of oligo(dT) at 5 $\mu$g/$\mu$l (Collaborative Research), 2 $\mu$l of each deoxynucleoside triphosphate at 25 mM (United States Biochemicals, Cleveland), and 4 $\mu$l of reverse transcriptase (Life Sciences, St. Petersburg, FL; 24 units/$\mu$l). After 40 min at 42°C the reaction was terminated by heating to 70°C for 10 min. To the reaction mixture was then added a solution containing 320 $\mu$l of RNase-free water, 80 $\mu$l of buffer [containing 0.1 M Tris·HCl (pH 7.5), 25 mM $MgCl_2$, 0.5 M KCl, bovine serum albumin at 0.25 mg/ml, and 50 mM dithiothreitol], 25 units of DNA polymerase I (Boehringer Mannheim), and 4 units of RNase H (Bethesda Research Laboratories). After 1 hr at 15°C and 1 hr at 22°C, 20 $\mu$l of 0.5 M EDTA (pH 8.0) was added, the reaction mixture was extracted with phenol, NaCl was added to 0.5 M, linear polyacrylamide (carrier; ref. 21) was added to 20 $\mu$g/ml, and the tube was filled with ethanol. After centrifugation for 2–3 min at 12,000 × *g*, the tube was removed, Vortex mixed to dislodge precipitate spread on the wall of the tube, and centrifuged again for 1 min.

Unpurified oligonucleotides having the sequence CTCTAAAG and CTTTAGAGCACA were dissolved in $H_2O$ at a concentration of 1 mg/ml. $MgSO_4$ was added to 10 mM, and the DNA was precipitated by adding 5 volumes of EtOH. The pellet was rinsed with 70% (vol/vol) EtOH and resuspended in TE buffer [10 mM Tris·HCl (ph 7.5)/0.5 mM EDTA] at a concentration of 1 mg/ml. The resuspended oligonucleotides (25 $\mu$l) were phosphorylated by the addition of 3 $\mu$l of buffer [containing 0.5 M Tris·HCl (pH 7.5), 10 mM ATP, 20 mM dithiothreitol, 10 mM spermidine, bovine serum albumin at 1 mg/ml, and 10 mM $MgCl_2$] and 20 units of polynucleotide kinase followed by incubation at 37°C overnight.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02988)

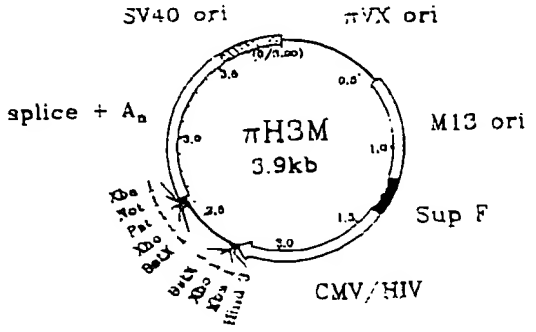


FIG. 1. Map of the πH3M vector. The direction of transcription is indicated by an arrow. Restriction endonuclease sites flanking the *Bst*XI cloning sites are shown. There are seven segments. Residues 1–587 are from the pBR322 origin (ori) of replication, residues 588–1182 are from the M13 origin, residues 1183–1384 are from the *supF* gene, residues 1385–2238 are from the chimeric cytomegalovirus/human immunodeficiency virus promoter, residues 2239–2647 are from the replaceable fragment, residues 2648–3547 are from plasmid pSV2 (splice and polyadenylylation signals), and residues 3548–3900 are from the simian virus 40 origin (SV40 ori). The complete nucleotide sequence is available from the authors.

The 12-mer (3 μl) and the 8-mer (2 μl) phosphorylated oligonucleotides were added to the cDNA prepared as above in a 300-μl reaction mixture containing 6 mM Tris·HCl (pH 7.5), 6 mM $MgCl_2$, 5 mM NaCl, bovine serum albumin at 0.35 mg/ml, 7 mM 2-mercaptoethanol, 0.1 mM ATP, 2 mM dithiothreitol, 1 mM spermidine, and 400 units of T4 DNA ligase (New England BioLabs) at 15°C overnight. EDTA (10 μl at 0.5 M) was added; and the reaction was phenol extracted, ethanol precipitated, resuspended in a volume of 100 μl of TE, and layered on a 5-ml gradient containing 5–20% (wt/vol) potassium acetate, 1 mM EDTA, and ethidium bromide at 1 μg/ml. The gradient was centrifuged for 3 hr at 50,000 rpm (SW55 rotor) and fractionated manually, collecting three ≈0.5-ml fractions followed by six ≈0.1-ml fractions in microcentrifuge tubes by means of a butterfly infusion set inserted just above the curve of the tube. Linear polyacrylamide was added to 20 μg/ml, the tubes were filled with ethanol, chilled, centrifuged, vortex mixed, and centrifuged again as above. The precipitate was washed with 70% (vol/vol) ethanol, dried, and resuspended in 10 μl. One microliter of the last six fractions was electrophoresed on a gel to determine which fractions to pool, and typically material <1 kilobase (kb) was discarded. Remaining fractions were pooled and ligated to the vector.

The complete sequence and derivation of the vector are available on request. The vector was prepared for cloning by digestion with *Bst*XI and by fractionation on 5–20% (wt/vol) potassium acetate gradients as described for the cDNA. The appropriate band was collected by syringe under 300-nm UV light and ethanol precipitated as above. cDNA and vector were titrated in test ligations. Usually 1–2 μg of purified vector was used for the cDNA from 4 μg of poly(A)⁺ RNA. The ligation reaction mixtures were composed as described for the adaptor addition above. The ligation reaction mixtures were transformed into MC1061/p3 cells made competent by an unpublished protocol available from Michael Scott (Department of Neurology, University of California, San Francisco) (personal communication). The transformation efficiency for supercoiled vector was $3–5 \times 10^6$ colonies per μg.

**Recovery and Characterization of the CD28 Clone.** Screening of the library was carried out as described by Seed and Aruffo (12), using purified antibody 9.3 (DuPont) at a concentration of 1 μg/ml of the antibody buffer. The methods used for COS cell transfection, radioimmunoprecipitation, RNA and DNA blot hybridization, and DNA sequencing were all as described (12).


FIG. 2. Nucleotide sequence of the CD28 cDNA. Nucleotide numbering is given in parentheses at right; amino acid numbering is above the sequence and not in parentheses. Locations of the potential sites for addition of asparagine-linked carbohydrate (CHO) are shown, as well as the predicted transmembrane (TM) sequence. The amino acid sequence is numbered from the projected cleavage site of the secretory signal sequence. The presumed polyadenylylation signal is shown boxed.

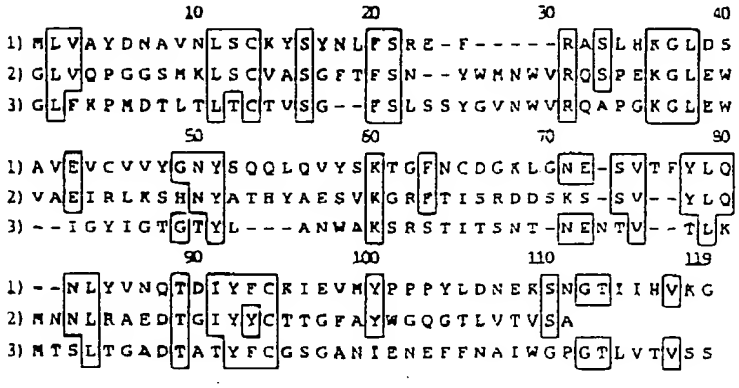

FIG. 3. Homology between the CD28 cDNA (lines 1), a mouse (lines 2), and a rabbit (lines 3) immunoglobulin heavy-chain variable region is shown boxed.

## RESULTS AND DISCUSSION

To isolate the CD28 cDNA, a large plasmid cDNA library was constructed in a high-efficiency expression vector containing a simian virus 40 origin of replication. A version of the vector, containing an M13 origin, is shown in Fig. 1. Three features of the vector make it particularly suitable for this use. (*i*) The eukaryotic transcription unit allows high-level expression in COS cells of coding sequences placed under its control. (*ii*) The small size and particular arrangement of sequences in the plasmid permit high-level replication in COS cells. (*iii*) The presence of two identical *Bst*XI sites in inverted orientation and separated by a short replaceable fragment allows the use of an efficient oligonucleotide-based strategy to promote cDNA insertion in the vector.

The *Bst*XI cleavage site, $CCAN_5/NTGG$ (where the slash denotes the cleavage site), creates a 4-base 3′ extension that varies from site to site. We created a vector in which two identical sites were placed in inverted orientation with respect to each other and separated by a short replaceable segment of DNA. Digestion with *Bst*XI followed by removal of the replaceable segment yields a vector molecule that is capable of ligating to fragments having the same ends as the replaceable segment but not to itself. In parallel, we attached to the cDNA synthetic oligonucleotides that give the same termini as the replaceable segment. The cDNA then cannot ligate to itself but can ligate to the vector. In this way, cDNA and vector are used as efficiently as possible. Tailing with terminal transferase achieves the same end but with less convenience and, in our hands, less overall efficiency. Moreover, homopolymer tracts located on the 5′ side of the cDNA inserts have been reported to inhibit expression *in vitro* and *in vivo* (15, 22, 23). Similar approaches based on the use of partially filled restriction sites to favor insertion of genomic DNAs (24) and cDNAs (16) have been reported. These approaches give 2- or 3-base complementary termini, which usually ligate less efficiently than the 4-base extensions reported here.

Although our cloning scheme does not result in a directional insertion of the cDNA, the ability to make large libraries easily, coupled with a powerful selection procedure, makes directional insertion unnecessary. In preliminary studies considerable effort was devoted to developing an efficient bidirectional transcription unit that would allow either orientation to be expressed at high levels; but it appears that this goal cannot be easily attained in COS cells because of mutual interference arising when complementary transcripts are formed. The library construction efficiencies we observe, between 0.5 and 2 × $10^6$ recombinants per $\mu$g of mRNA with <1% background and an insert size >1 kb, compare favorably with those described for phage vectors λgt10 (7.5 × $10^5$ recombinants per $\mu$g of mRNA) and λgt11 (1.5 × $10^6$ recombinants per $\mu$g of mRNA) (18); but the resulting clones are more convenient to manipulate.

Surface antigen cDNAs can be isolated from these libraries using an antibody-enrichment method (12). In this method, spheroplast fusion is used to introduce the library into COS cells, where it replicates and expresses its inserts. The cells are harvested by detaching without trypsin, treated with monoclonal antibodies specific for the surface antigens desired, and distributed in dishes coated with affinity-purified antibody to mouse immunoglobulins (12). Cells expressing surface antigen adhere, and the remaining cells can be washed away. [This general method of cell selection is known as "panning" (25)]. From the adherent cells, a "Hirt" fraction is prepared (26), and the resulting DNA is transformed back into *Escherichia coli* for further rounds of fusion and selection. Typically, after two rounds of selection with monoclonal antibodies recognizing various surface antigens, a single round of selection is performed with a single antibody or a pool of antibodies recognizing the same antigen (unpublished results).

**Isolation of a CD28 cDNA.** The CD28 cDNA was isolated from a library of ≈3 × $10^5$ recombinants prepared from cDNA from 0.8 $\mu$g of poly(A)$^+$ RNA. The library was screened for CD28 (and other surface antigen) cDNA clones by the method outlined above (12). After the third transfection, COS cells were panned with the 9.3 antibody alone. A Hirt supernatant was prepared from the adherent cells and transformed into *E. coli*. Plasmid DNA was isolated from eight colonies and transfected individually into COS cell cultures. The presence of the CD28 antigen was detected in three of eight transfected cultures by indirect immunofluorescence. All three plasmid DNAs contained an insert of ≈1.5 kb.

**cDNA Sequence Analysis.** The CD28 cDNA encodes a long

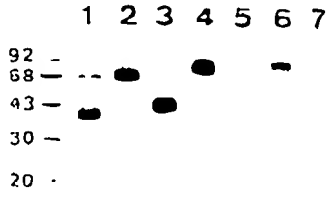


FIG. 4. Immunoprecipitation of CD28 antigen synthesized by T cells and transfected COS cells. Lanes: 1, 3, and 5, CD28 antigen from COS cells, activated T cells, and HPB-ALL cells, under reducing conditions; 2, 4, and 6, CD28 from COS cells, activated T cells, and HPB-ALL cells, under nonreducing conditions; 7, CD28-transfected COS cells treated with anti-CD4 antibody.

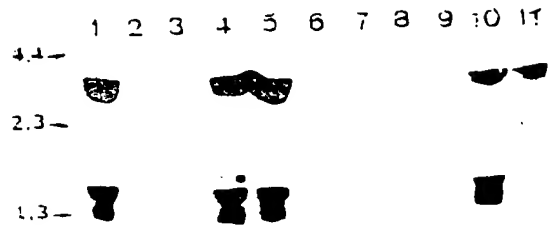


FIG. 5. RNA blot analysis of CD28-related transcripts. RNA sources were as follows. Lanes: 1, thymocytes; 2, U-937; 3, HuT 78; 4, T blast; 5, HPB-ALL; 6, Jurkat J3R7; 7, Namalwa; 8, MOLT-4; 9, HSB-2; 10, PEER; and 11, senescent (rested) T cells. RNA sizes (in kb) are from poly(A)$^+$ standards (Bethesda Research Laboratories).

open reading frame of 220 residues having the typical features of an integral membrane protein (Fig. 2). Removal of a predicted (27) N-terminal signal sequence gives a mature protein of 202 residues comprising an extracellular domain with five potential N-linked glycosylation sites (Asn-Xaa-Ser/Thr), a 27-amino acid hydrophobic membrane-spanning domain, and a 41-amino acid cytoplasmic domain. Comparison of the amino acid sequence of CD28 with the National Biomedical Research Foundation Database† revealed substantial homology with mouse and rabbit immunoglobulin heavy-chain variable regions over a domain spanning almost the entire extracellular portion of CD28 (Fig. 3). Within this domain two cysteine residues in the homology blocks Leu-(Ser or Thr)-Cys and Tyr-(Tyr or Phe)-Cys are shared by CD28, CD4, CD8, immunoglobulin heavy- and light-chain variable sequences, and related molecules with approximately the same spacing (28–30).

**CD28 cDNA Directs the Production of a Homodimer in Transfected COS Cells.** Immunoprecipitation of CD28 antigen from transfected COS cells was carried out using the monoclonal antibody 9.3 (10). The material obtained from COS cells migrated with a molecular mass of 74 kDa under nonreducing conditions and 39 kDa under reducing conditions (Fig. 4), a pattern consistent with homodimer formation. Under the same conditions activated T cells give bands with molecular masses of 87 and 44 kDa, and HPB-ALL cells give bands of 92 and 50 kDa, under nonreducing and reducing conditions, respectively. The variation in molecular mass of the material obtained from various cell types likely arises as a result of differing glycosylation patterns characteristic of each type. We have observed similar results with other leukocyte surface antigens (B.S., unpublished results and ref. 12). The nucleotide sequence of the CD28 cDNA predicts a mature protein with molecular mass of 23 kDa, much smaller than observed in these experiments, and probably attributable to utilization of the five N-linked glycosylation sites predicted by the amino acid sequence.

**RNA Blot Analysis.** Equal amounts of total RNA prepared from cell types expressing or lacking CD28 were subjected to RNA blot analysis as described (12). Four bands (Fig. 5) with molecular sizes of 3.7, 3.5, 1.5, and 1.3 kb were visible in lanes containing RNA from thymocytes, T blasts, senescent T cells, and the T-cell leukemia cell lines PEER and HPB-ALL. No bands were detected in lanes containing RNA

†Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 10.0.



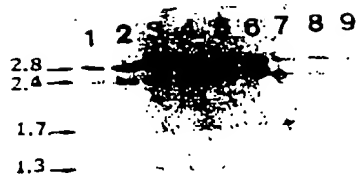


FIG. 6. Genomic DNA blot analysis. Genomic DNA (20 μg) digested with *Dra* I was electrophoresed through [illegible] agarose, blotted, and hybridized to a CD28 probe. DNA [illegible] were prepared from the following cells. Lanes: 1, HuT 78 cells; 2, PEER cells; 3, MOLT-4 cells; 4, HSB-2 cells; 5, placenta; 6, Jurkat J3R7 cells; 7, total peripheral blood; 8, HPB-ALL cells; and 9, T blasts.

prepared from the cell lines U-937 (histiocytic leukemia), HuT 78 (adult T-cell leukemia), Jurkat (T-cell leukemia), Namalwa (Burkitt lymphoma), MOLT-4, and HSB-2, all of which do not express CD28. We presume that the 3.5-kb transcript corresponds to the isolated cDNA and that the 3.7- and 3.5-kb species reflect incomplete splicing or alternative polyadenylylation site utilization.

**The CD28 Gene Is Not Rearranged.** DNA blot analysis (12) of genomic DNA from placenta, peripheral blood lymphocytes, T cells, Hela cells, or the tumor lines used in the RNA blot analysis above showed identical *Dra* I digest patterns (Fig. 6), indicating that rearrangement is not involved in the normal expression of the CD28 gene during development. Similarly, no gross genomic rearrangement underlies the failure of the examined T-cell tumor lines to express CD28 antigen. We infer from the *Dra* I fragment pattern that the CD28 gene contains at least two introns.

This work was supported by a grant to the Massachusetts General Hospital by Hoechst.

1. Weiss, A., Wiskocil, R. L. & Stobo, J. D. (1984) *J. Immunol.* 133, 123–128.
2. Holter, W., Fischer, G. F., Mantle, D., Stockinger, H. & Knapp, W. (1986) *J. Exp. Med.* 163, 654–664.
3. Hara, T., Fu, S. M. & Hansen, J. A. (1985) *J. Exp. Med.* 161, 1513–1524.
4. Van Wauwe, J. P., De Mey, J. R. & Goossens, J. G. (1980) *J. Immunol.* 124, 2708–2713.
5. Brenner, M. B., Trowbridge, I. S. & Strominger, J. L. (1985) *Cell* 40, 183–190.
6. Katz, S. P., Kierszenbaum, F. & Waksman, B. H. (1978) *J. Immunol.* 121, 2286–2291.
7. Ledbetter, J. A., Martin, P. J., Spooner, C. E., Wofsy, D., Tsu, T. T., Beatty, P. G. & Gladstone, P. (1985) *J. Immunol.* 135, 2331–2336.
8. Williams, J. M., Deloria, D., Hansen, J. A., Dinarello, C. A., Loertscher, R., Shapiro, A. M. & Strom, T. (1985) *J. Immunol.* 135, 2249–2255.
9. Dower, S. K., Kronheim, S. R., March, C. J., Conlon, P. J., Hopp, T. P., Gillis, S. & Urdal, D. L. (1985) *J. Exp. Med.* 162, 501–515.
10. Hansen, J. A., Martin, P. J. & Nowinski, R. C. (1980) *Immunogenetics* 10, 247–260.
11. Ledbetter, J. A., Parsons, M., Martin, P. J., Hansen, J. A., Rabinovitch, P. S. & June, C. H. (1986) *J. Immunol.* 137, 3299–3305.
12. Seed, B. & Aruffo, A. (1987) *Proc. Natl. Acad. Sci. USA* 84, 3365–3369.
13. Wong, G. G., Witek, J., Temple, P. A., Wilkens, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, E. C., Shoemaker, C., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A. & Clark, S. C. (1985) *Science* 228, 810–815.
14. Lee, F., Yokota, T., Otsuka, T., Meyerson, P., Villaret, D., Coffman, R., Mosmann, R., Rennick, D., Roehm, N., Smith, C., Zlotnik, A. & Arai, K. I. (1986) *Proc. Natl. Acad. Sci. USA* 83, 2061–2065.

Yokota, T., Otsuka, T., Mosmann, T., Banchereau, J., DeFrance, T., Blanchard, D., De Vries, J. E., Lee, F. & Arai, K. (1986) *Proc. Natl. Acad. Sci. USA* 83, 5894–5898.
Yang, Y., Ciarletta, A. B., Temple, P. A., Chung, M. P., Kovacic, S., Witek-Giannotti, J. S., Leary, A. C., Kriz, R., Donahue, R. E., Wong, G. G. & Clark, S. C. (1986) *Cell* 47, 3–10.
Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* 2, 161–170.
Huynh, T., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 18, 5294–5299.
Gubler, U. & Hoffman, B. J. (1983) *Gene* 25, 263–269.
Strauss, F. & Varshavsky, A. (1984) *Cell* 37, 889–901.
Galili, G., Kawata, E. E., Cuellar, R. E., Smith, L. D. & Larkins, B. A. (1986) *Nucleic Acids Res.* 14, 1511–1524.
Riedel, H., Kondor-Koch, C. & Garoff, H. (1985) *EMBO J.* 3, 1477–1483.
Zabarovsky, E. R. & Allikmets, R. L. (1986) *Gene* 42, 119–123.
Wysocki, L. J. & Sato, V. L. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2844–2848.
Hirt, B. (1967) *J. Mol. Biol.* 26, 365–369.
von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683–4690.
Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L. & Axel, R. (1985) *Cell* 42, 93–104.
Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L. & Axel, R. (1985) *Cell* 40, 237–246.
Amzel, L. M. & Poljak, R. J. (1979) *Annu. Rev. Biochem.* 48, 961–997.

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1997, p. 6472–6480
0270-7306/97/$04.00+0

Vol. 17, No. 11

# O Glycosylation of an Sp1-Derived Peptide Blocks Known Sp1 Protein Interactions

MARK D. ROOS,[1] KAIHONG SU,[1] JOHN R. BAKER,[2] AND JEFFREY E. KUDLOW[3]*

*Departments of Medicine/Endocrinology,[3] Cell Biology,[1] and Biochemistry and Molecular Genetics,[2] University of Alabama at Birmingham, Birmingham, Alabama 35294*

Received 24 March 1997/Returned for modification 13 May 1997/Accepted 27 August 1997

**The O-linked *N*-acetylglucosamine (*O*-GlcNAc) modification of proteins is dynamic and abundant in the nucleus and cytosol. Several transcription factors, including Sp1, have been shown to contain this modification; however, the functional role of *O*-GlcNAc in these proteins has not been determined. In this paper we describe the use of the previously characterized glutamine-rich transactivation domain of Sp1 (B-c) as a model to investigate the role of *O*-GlcNAc in Sp1's transcriptionally relevant protein-to-protein interactions with the TATA-binding-protein-associated factor (TAF110) and holo-Sp1. When the model Sp1 peptide was overexpressed in primate cells, this 97-amino-acid domain of Sp1 was found to contain a dominant *O*-GlcNAc residue at high stoichiometry, which allowed the mapping and mutagenesis of this glycosylation site. In vitro interaction studies between this segment of Sp1 and *Drosophila* TAF110 or holo-Sp1 indicate that the *O*-GlcNAc modification functions to inhibit the largely hydrophobic interactions between these proteins. In HeLa cells, the mutation at the mapped glycosylation site was permissive for transcriptional activation. We propose the hypothesis that the removal of *O*-GlcNAc from an interaction domain can be a signal for protein association. *O*-GlcNAc may thereby prevent untimely and ectopic interactions.**

Many cytosolic and nuclear proteins (11, 13, 14, 16, 17) are covalently modified by the addition of monomeric O-linked *N*-acetylglucosamine (*O*-GlcNAc) groups. Furthermore, this modification has been found to undergo dynamic changes, often in a signal-dependent manner (21). Many proteins that form multimeric complexes have been shown to be glycosylated with *O*-GlcNAc (*O*-GlcNAcylated), and this modification has been compared to phosphorylation (13) with respect to controlling protein function. The aspects of proteins speculated to be altered by modification with *O*-GlcNAc (*O*-GlcNAcylation) include protein-protein interactions (13, 14, 17), protein stability (12, 29), and subcellular localization (29). The observation that many transcription factors are *O*-GlcNAcylated suggested that this modification may play a role in the control of transcription (19). To date, there has not been any direct evidence to implicate *O*-GlcNAcylation in any of these roles.

We chose Sp1 as a model to better define the role of *O*-GlcNAcylation. Sp1 is an ubiquitous transcription factor that plays a particularly vital role in the regulation of transcription from TATA-less promoters that commonly encode housekeeping genes (25). It is a well-characterized protein composed of 778 amino acids. The amino-terminal portion of the molecule contains two glutamine-rich domains, each of which is associated with serine-threonine-rich regions (20). These domains are involved in transcriptional activation. The carboxy-terminal region of the molecule contains the zinc-finger DNA-recognition domain. Sp1 is known to be phosphorylated upon binding DNA (18), and each molecule of Sp1 is thought to bear at least nine *O*-GlcNAc residues (19). More-detailed analysis of the glutamine-rich activation domains indicates that they are involved in both the homomultimerization of Sp1 (24) and the interaction with TFIID (8, 15). By a yeast two-hybrid assay, a small region of Sp1 (amino acids 424 to 542, the B-c domain) has been shown to interact with the TFIID protein TATA-binding-protein-associated factor 110 (TAF110) (8), and by using *Drosophila* cells, this same region of Sp1 has been shown to be involved in self-association (24). Mutagenesis of the Sp1 B-c domain has indicated that the interaction between the B-c domain and TAF110 involves the glutamine-rich hydrophobic patch present in this domain (8) of Sp1 and conserved in other transcription factor activation domains such as CREB (7) and VP16 (acidic and hydrophobic) (5).

Like Sp1 (19), several transcription factors have been shown to be *O*-GlcNAcylated. With the serum response factor (SRF) (26) and c-Myc (1), where the sites of modification were determined, modified residues were mapped to the transcriptional activation domains of these proteins. These findings raised the prospect that the *O*-GlcNAc modification alters the ability of these factors to regulate transcription. Indeed, indirect evidence from an in vitro transcription reaction indicated that Sp1-dependent transcription could be blocked by the GlcNAc-binding lectin, wheat germ agglutinin (19). We were further encouraged to pursue this hypothesis by the observation that the Sp1 B-c domain contains a sequence that is homologous to the sequence surrounding the *O*-GlcNAcylation site in the activation domain of SRF. The prior characterization of the interaction of the Sp1 B-c domain with holo-Sp1 and the coactivator, TAF110, provided the opportunity to obtain more direct evidence as to whether these transcriptionally relevant interactions involving the B-c domain could be altered by the *O*-GlcNAc modification. To this end, we created a model system based on the Sp1 B-c domain in which we could determine if this domain is *O*-GlcNAcylated and in which we could map and mutate the site(s) of glycosylation. By site-directed mutagenesis or the use of different expression systems, we could alter the degree of glycosylation of this Sp1 peptide and test the effect of this altered glycosylation on the known protein-protein interactions in which Sp1 engages. Using this Sp1 model in an in vitro protein interaction assay, we found that *O*-GlcNAcylation of a single serine residue, near

* Corresponding author. Mailing address: Department of Medicine/Endocrinology, University of Alabama at Birmingham, 1808 7th Ave. South, Room 756, Birmingham, AL 35294. Phone: (205) 934-4116. Fax: (205) 934-4389. E-mail: kudlow@endo.dom.uab.edu.

the glutamine-rich hydrophobic patch contained in this Sp1-derived peptide, markedly diminished interactions with holo-Sp1 and TAF110. This finding suggests that one function of *O*-GlcNAcylation is to inhibit the otherwise strong hydrophobic interactions that occur between transcription factors. We postulate that inhibition of these hydrophobic interactions must occur to prevent untimely, ectopic, and nonspecific protein-protein interactions between transcription factors and TAFs prior to their proper assembly on the cognate DNA template.

## MATERIALS AND METHODS

**Materials.** The following materials were obtained from the indicated suppliers: a $C_{18}$ high-performance liquid chromatography (HPLC) column (catalog no. 86-200-C5; Rainin, Milpitas, Calif.); glutathione-Sepharose (Pharmacia, Piscataway, N.J.); proline-specific endopeptidase (Seikagaku); acetonitrile (EM Science, Gibnstown, N.J.); trifluoroacetic acid (TFA; Pierce, Rockford, Ill.); a manual Edman peptide sequencing kit (Millipore, Bedford, Mass.); UDP-[6-$^3$H]galactose (38 Ci/mmol; Amersham, Arlington Heights, Ill.); and thrombin, GlcNAc, and bovine milk galactosyltransferase (Sigma, St. Louis, Mo.). The galactosyltransferase was pregalactosylated as described previously (33).

**Cell culture.** BSC40 cells were grown in Dulbecco modified Eagle medium with 10% NCS (Gibco BRL, Grand Island, N.Y.), 100 μg of penicillin/ml, and 50 μg of gentamicin/ml at 37°C in a humidified incubator with 7.5% $CO_2$.

**Expression of fusion proteins in vaccinia virus and *Escherichia coli*.** The cDNA for glutathione *S*-transferase (GST) was PCR amplified with Pharmacia's pGEX vector as a template such that it could be cloned into the pTM3 vector (23) between the *Nco*I and *Eco*RI sites. The cDNA that encodes the 97 amino acids (424 to 521) of Sp1 (SpE) was PCR amplified, sequenced, and cloned into this GST-modified pTM3 vector. The serine residue in SpE corresponding to serine 484 in Sp1 was converted to an alanine by site-directed mutagenesis to generate the SpS peptide. Recombinant vaccinia viruses that allowed expression of the GST-SpE and the GST-SpS fusion proteins were generated (22). The GST-SpE fusion protein and GST were expressed in *E. coli* with the pGEX vector and purified on glutathione-Sepharose according to Pharmacia's pGEX protocol.

**Purification of SpE and SpS.** After infection of BSC40 cells with recombinant vaccinia virus for 36 h, an extract was prepared by freezing and thawing the cells twice in extraction buffer containing 20 mM Tris (pH 7.5), 0.5 M NaCl, 0.5 M GlcNAc, 0.5 mM EDTA, 0.5 mM $MgCl_2$, 1 mM dithiothreitol, 0.2 mM phenylmethylsulfonyl fluoride, and 20% glycerol. The supernatant was collected after centrifugation. Glutathione-Sepharose was added to the extract for 30 min. The beads were collected and washed three times with the extraction buffer minus the glycerol and GlcNAc. GlcNAc residues were labeled with [$^3$H]galactose as described previously (33). The GST peptides were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining. These peptides were at least 90% pure at this stage and were of the predicted molecular weight. For HPLC purification of peptides, the affinity beads were washed three times with a buffer containing 20 mM Tris (pH 7.5), 150 mM NaCl, and 2.5 mM $CaCl_2$. Then SpE or SpS was cleaved from GST with 4 U of thrombin per mg of fusion protein.

**RP-HPLC.** Peptides were isolated by reverse-phase HPLC (RP-HPLC) on a Beckman System Gold HPLC system equipped with a Rainin $C_{18}$ column. Peptides were bound to the column in aqueous 0.01% TFA and eluted with increasing concentrations of acetonitrile as indicated in Fig. 3. Elution of peptides was monitored by UV absorption at 214 nm, and the [$^3$H]galactose-labeled glycopeptides were detected by liquid scintillation spectrometry of a small aliquot of each fraction collected.

**Chemical and enzymatic digestion of SpE.** After thrombin digestion of the fusion protein and HPLC purification of the SpE peptide, 100 μg of the peptide was then cleaved with cyanogen bromide (CNBr) at 15 mg/ml in 0.01% TFA and 2 M guanidine HCl for 24 h at 25°C. After HPLC purification, 50 μg of the labeled CNBr-cleaved peptide was then digested with 0.2 U of a proline-specific endopeptidase in a 50 mM Tris (pH 7.5) buffer for 2 h at 25°C.

**Gas-phase sequencing and manual Edman degradation.** The primary sequences of HPLC-purified peptides were determined by gas-phase automated Edman degradation on a Beckman model PI 2090 peptide microsequencer equipped with an on-line PTH amino acid analyzer. The site of glycosylation was determined by manual Edman degradation with modifications as described by Kelly et al. (22). In short, the peptide was immobilized on an aryl amine membrane. Single amino acids were then removed by Edman degradation chemistry. The released amino acid derivative from each cycle was then collected, dried, neutralized, and assayed by scintillation spectrometry to determine the cycle at which the [$^3$H]galactose-labeled amino acid was released.

**Mass spectroscopy of glycopeptides.** The electrospray mass spectra of HPLC-purified peptides were obtained on a Sciex (Concorde, Ontario, Canada) API-III triple quadrapole mass spectrometer equipped with an atmospheric-pressure ion source. The peptide sample was injected onto a capillary $C_{18}$ RP-HPLC column equilibrated in 0.1% formic acid and eluted into the mass spectrometer with acetonitrile at a flow rate of 4 μl/min. Positive-ion mass spectra were acquired at an orifice potential of 70 V (to scanning the first quadrapole (Q1) over a mass-to-charge ratio (*m/z*) range of 400 to 2,000 at a rate of 6 s per scan. Data were acquired and processed with Sciex API-MacSpec software, version 3.22.

Matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) mass spectroscopy of the Sp1 peptides was carried out on a Perspective Biosystems (Framingham, Mass.) Voyager Elite MALDI-TOF mass spectrometer. Samples were mixed with a saturated solution of α-cyano-4-hydroxy-cinnamic acid in a water-acetonitrile (50:50) mixture acidified with 0.1% TFA. A 1-μl aliquot of the sample was spotted onto the gold plate target. Ionization of the sample was accomplished with a nitrogen laser operated at 337 nm. A delayed-extraction method was used in the determination of molecular mass. Measurement of ion flight times through the drift region of the mass spectrometer was carried out with a Tektronix (Beaverton, Oreg.) model TDS784A oscilloscope. The instrument was calibrated with external molecular weight standards.

**In vitro protein interaction studies.** Equal quantities of the GST fusion proteins were immobilized on glutathione-Sepharose affinity beads and washed extensively with the extraction buffer containing 0.05% Nonidet P-40 but no glycerol or GlcNAc (binding buffer). The cDNA for *Drosophila* TAF110 was kindly provided by R. Tjian (15). The cDNA for Sp1 was kindly provided by S. Kadonaga. These cDNAs were placed in plasmids downstream of the T7 promoter. The Promega TNT kit for wheat germ extract was used as directed by the manufacturer to synthetically label Sp1 and TAF110 with [$^{35}$S]methionine. The $^{35}$S-labeled Sp1 or TAF110 in 10 μl of the wheat germ lysate was added to 30 μl of a 50% slurry of glutathione affinity beads to which equal amounts of GST or GST fusion proteins (~1 μg) were bound. After a 2-h incubation at 4°C, the beads were washed extensively with the binding buffer and then boiled in SDS-PAGE sample buffer. The proteins bound to the beads were separated by SDS-PAGE. The $^{35}$S-labeled proteins that had bound to the beads were observed by fluorography, while the unlabeled GST fusion proteins were observed following Coomassie blue staining of the gel.

**Transient transfection of HeLa cells.** HeLa cells ($7 \times 10^6$) were used for each transfection. Cells were trypsinized, washed two times with cold phosphate-buffered saline, resuspended in 0.4 ml of Dulbecco modified Eagle medium plus 10% newborn calf serum. Twenty micrograms of a Gal4-dependent luciferase reporter plasmid, 0.5 μg of an expression plasmid encoding the Gal4 fusion proteins, and 20 μg of a cytomegalovirus β-galactosidase (30) plasmid were added to the cells. The Gal4 expression plasmids contained the simian virus 40 promoter upstream of the cDNA encoding the first 94 amino acids of Gal4 fused in frame to the cDNA encoding SpE or SpS. Cells were electroporated at 350 V and 500 μF in a Gene Pulser (Bio-Rad, Richmond, Calif.) and plated at a density of $10^6$ cells per well onto six-well plates (Fisher Scientific, Atlanta, Ga.). Cells were harvested and assayed for luciferase and β-galactosidase activities 24 h posttransfection (30). The luciferase activity was normalized to the β-galactosidase activity to control for variations in transfection efficiency. All transfections were done in triplicate.

## RESULTS

**The Sp1 model peptide, SpE, is glycosylated.** Previous characterization of Sp1 (8, 24) indicated that the carboxy-terminal region of the second transcriptional activation domain (Fig. 1), the B-c domain, could confer protein-protein interactions with both holo-Sp1 and TAF110. We also observed that this region of Sp1 contains a sequence resembling the sequence in the SRF transcriptional activation domain to which an *O*-GlcNAc site was mapped (26). As shown in Fig. 1, 7 of 9 amino acids in the SRF sequence match Sp1. Therefore, a recombinant vaccinia virus that would direct the expression in primate cells of a GST-SpE fusion protein (amino acids 424 to 521 of Sp1) containing most of the B-c domain was generated. The recombinant baculovirus system that expresses proteins in an insect cell line has been used to identify *O*-GlcNAc sites in other proteins (1, 26). The GST-SpE fusion protein was extracted from a whole-cell lysate of recombinant vaccinia virus-infected BSC40 (monkey kidney) cells under conditions that minimized hexosaminidase activity (1) and partially purified by affinity chromatography on glutathione-Sepharose beads. As shown in Fig. 2, a thrombin cleavage site was engineered between GST and SpE to allow the separation of the Sp1 peptide from GST. GST-SpE, immobilized on glutathione-Sepharose beads, was labeled at *O*-GlcNAc residues with tritiated galactose by using galactosyltransferase (33). The labeled fusion protein (Fig. 2, lane 1) was digested from GST with thrombin (lane 2) and released into the supernatant (lane 3). These proteins were

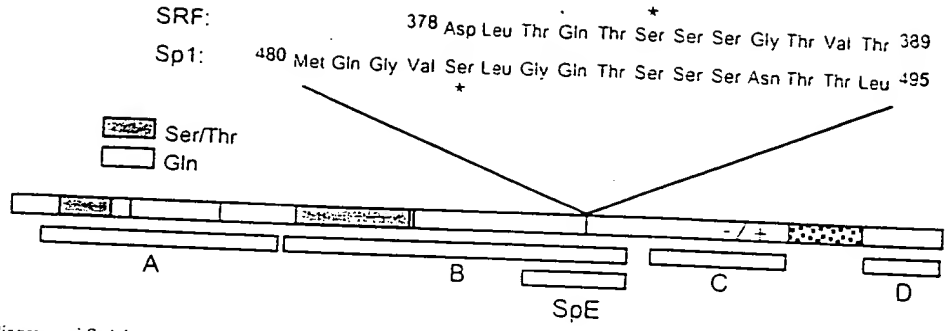


FIG. 1. Schematic diagram of Sp1 indicating the previously defined domains. The position of the SpE peptide is shown relative to that of Sp1. The amino acid sequence surrounding the glycosylation site (*) of GST-SpE is shown. Also, the homologous sequence of SRF is shown with its site of glycosylation (*).

analyzed by SDS-PAGE and fluorography. The label was incorporated into the GST-SpE fusion protein and the 12-kDa SpE released by thrombin, but the label was not incorporated into the 26-kDa GST. GST expressed by itself was not labeled with [$^3$H]galactose (data not shown), indicating that it is not a substrate for *O*-GlcNAc transferase in this system. Furthermore, no *O*-GlcNAc could be detected in the Sp1 peptide corresponding to amino acids 1 to 81 when it was expressed as a GST-fusion protein in the vaccinia virus system (data not shown). Thus, the *O*-GlcNAc transferase in vaccinia virus-infected primate cells appears to recognize the SpE moiety in the fusion protein with specificity.

**HPLC purification and characterization of SpE.** In order to use the *O*-GlcNAcylated SpE peptide as a model, the degree of glycosylation and the site of glycosylation were determined. The [$^3$H]galactose-labeled SpE that had been digested from GST was purified by HPLC. The identity of the SpE peptide was confirmed by MALDI-TOF mass spectroscopy (see below) and electrospray mass spectrometry (ESMS). Two molecular mass species were identified, one having a molecular mass corresponding to the peptide backbone alone and the other corresponding to the peptide plus 203 mass units. This extra 203 mass units is attributable to a single *O*-GlcNAc group attached to the peptide. This observation implies that either SpE has a single glycosylation site or it has two or more sites that are modified in a mutually exclusive manner. No peptide signal was detected with a mass corresponding to that of the *O*-GlcNAc peptide plus a galactose, suggesting that the conditions for galactose labeling resulted in markedly substoichiometric labeling. These data, and the low stoichiometry of [$^3$H]galactose labeling (data not shown), suggested that the glycosylated SpE peptide is a poor substrate for galactosyltransferase.

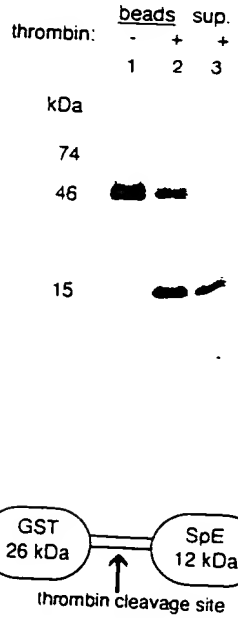


FIG. 2. The Sp1 peptide, SpE, contains GlcNAc. Vaccinia virus-expressed GST-SpE, immobilized on glutathione-Sepharose beads, was labeled with [$^3$H]galactose by using galactosyltransferase (lane 1) and then treated with thrombin (lanes 2 and 3). The proteins bound to the beads before (lane 1) and after (lane 2) thrombin digestion, and the proteins released from the beads (lane 3) were separated by SDS-PAGE. The [$^3$H]galactose ([$^3$H]Gal)-labeled proteins were observed by fluorography. Fluorography indicated that the galactose label was attached to the SpE peptide (12 kDa) and not GST (26 kDa).

To determine the amino acid(s) in SpE that is modified by *O*-GlcNAc, 100 μg of the peptide was fragmented. Because the purified SpE peptide was resistant to proteolysis and contained two methionine residues, the peptide was cleaved with CNBr in 0.01% TFA and 2 M guanidine HCl. The resulting peptides were separated by HPLC and identified by ESMS. As shown in Fig. 3, we generated predominantly four peptides whose molecular masses corresponded to those of the peptide sequences labeled A, B, D, and E in the figure. The [$^3$H]galactose-labeled glycopeptide eluted in fractions 18 through 20 and had a mass corresponding to that of peptide B (Fig. 3). No $^3$H label was associated with fractions that did not contain peptide B, and ESMS failed to detect an *O*-GlcNAc mass associated with peptides A and D. The identity of the labeled glycopeptide was further confirmed by automated amino acid sequencing.

Because peptides generated by CNBr cleavage cannot be covalently attached to the membrane used in manual Edman degradation, approximately 50 μg of glycopeptide B was further fragmented with 0.2 U of a proline-specific endopeptidase to liberate free carboxy termini. After HPLC purification of the $^3$H-labeled glycopeptide, it was covalently linked to a membrane and a single amino acid was released from the N terminus of the peptide at each cycle of a manual Edman degradation reaction. The labeled peptide was also sequenced with an automated sequencer. Figure 4 shows the amino acid sequence of the peptide as determined by gas-phase sequencing as well as the counts per minute released at each cycle of the manual Edman degradation reaction. Greater than 90% of the $^3$H label on the peptide was released from the membrane on the fourth cycle, indicating that the predominant galactose-labeled *O*-GlcNAcylation site in SpE corresponds to serine 484 in

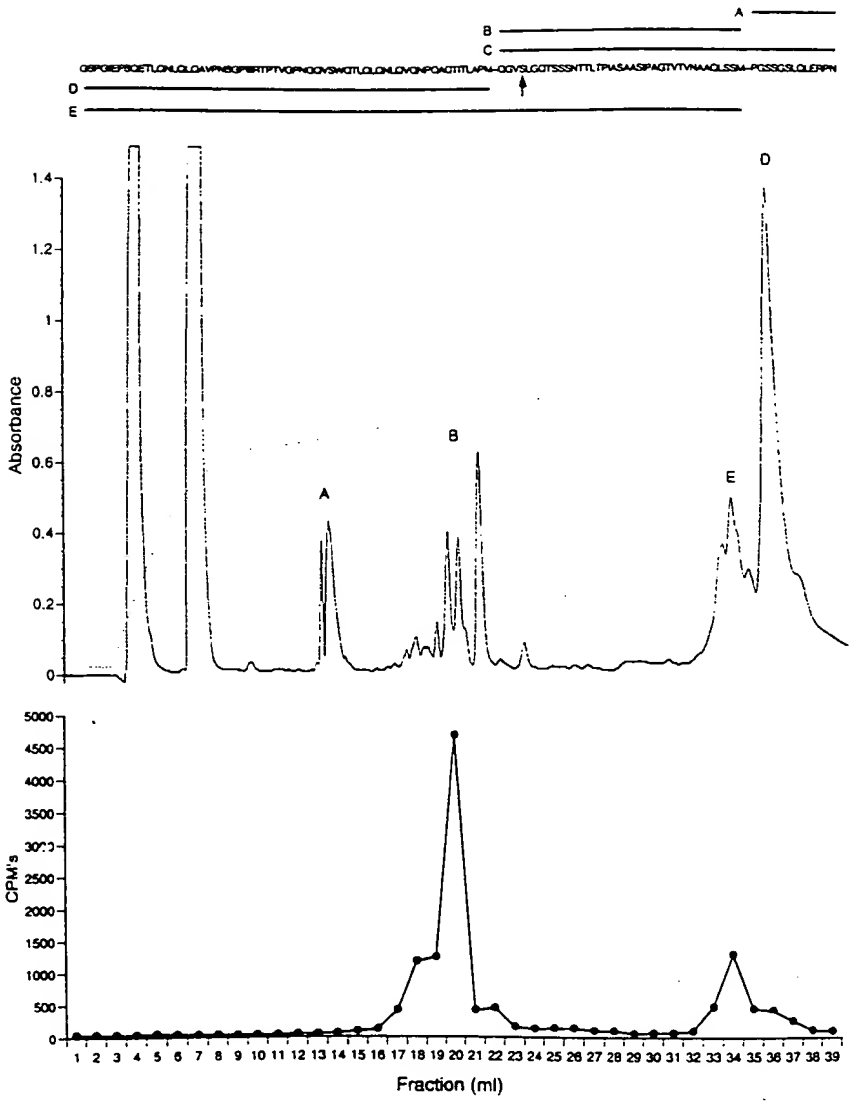


FIG. 3. HPLC purification of the glycopeptides derived from SpE. The SpE peptide was labeled with [$^3$H]galactose and purified by HPLC. The purified peptide was solubilized in guanidine hydrochloride and fragmented with CNBr. The CNBr fragments were applied to a $C_{18}$ RP-HPLC column and eluted with increasing concentrations of acetonitrile (the elution gradient is shown as a dotted line). Shown above the UV-absorbance profile are the predicted CNBr fragments of SpE, labeled A through E. The eluted peptides were identified by ESMS and are indicated on the UV profile at the positions of the corresponding absorbance peaks. One-milliliter fractions were collected, and the levels of radioactivity in these fractions were determined and are shown in the lower graph. Only fractions containing peptide B contained $^3$H, indicating that peptide B harbors the *O*-GlcNAc-modified residue. The arrow indicates the O-glycosylation site.

native Sp1. Of note, the homologous serine in SRF that bears the *O*-GlcNAc modification corresponds to serine 489 in Sp1 (Fig. 1). Thus, it appears that Sp1 and SRF differ in the precise serines recognized by the *O*-GlcNAc transferase. Nevertheless, these modified sites are close to each other and both occur in the transcriptional activation domain.

**Characterization of the peptide mutated at the *O*-GlcNAc site (SpS).** The serine residue that was determined to be the dominant site of *O*-GlcNAcylation in SpE was mutated to an alanine residue by site-directed mutagenesis to generate a peptide termed SpS. The GST-SpS fusion protein was generated in parallel with GST-SpE with the vaccinia virus expression system. Both the SpE and SpS peptides were purified by HPLC and both were subjected to an amino acid analysis to determine the degrees of glycosylation of these peptides. No glucosamine was detected in the SpS peptide. By contrast, SpE contained 0.7 mol of glucosamine per mol of peptide.

These results were confirmed qualitatively by MALDI-TOF mass spectroscopy and ESMS (Fig. 5). The MALDI analysis (Fig. 5A) of SpE expressed in vaccinia virus revealed the presence of two major species having molecular masses of 11,916.3 and 12,118.4 atomic mass units, which differ by 202.1 atomic mass units. This difference in mass is consistent with the presence of a single *O*-GlcNAc residue (203 mass units) on the larger peptide. The same peptide expressed in *E. coli* had a mass determined to be 11,919.5 atomic mass units. Based on the known amino acid sequence of SpE, the predicted molecular mass is 11,919.4 atomic mass units, in close agreement with these determinations. This result indicates that the only difference between the vaccinia virus- and bacterium-expressed

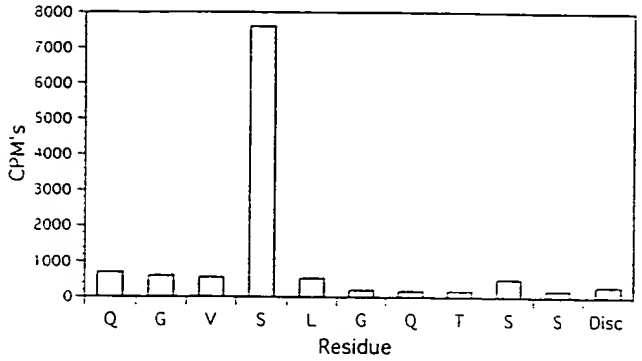


FIG. 4. Manual Edman degradation was performed on the fragment of SpE containing the [$^3$H]galactose-labeled *O*-GlcNAc. The HPLC-purified peptide was identified by automated N-terminal sequencing and ESMS. The [$^3$H]galactose-labeled SpE fragment was immobilized on a membrane disc, and the radioactivity eluted from the membrane-bound peptide was determined for each Edman cycle. The graph shows the radioactivity eluted at each cycle and the corresponding amino acids, as determined by automated peptide sequencing of the same peptide. The majority of the $^3$H was recovered in cycle 4, corresponding to serine 484 of Sp1.

SpE peptides is the presence, in the vaccinia virus material, of a peptide about 203 mass units larger than that in the bacterial product. This result implies that the vaccinia virus SpE does not contain any covalent modification, other than the addition of *O*-GlcNAc, that would result in a molecular mass shift. MALDI analysis of SpS expressed in vaccinia virus yielded a molecular mass of 11,899.2 atomic mass units, which is appropriately 16 atomic mass units smaller than the mass of bacterially expressed SpE, predicted to result from the serine-to-alanine mutation. In all of the samples, a small peak corresponding to an additional molecular mass of about 210 atomic mass units is evident. This species likely represents a chemical adduct resulting from the interaction of the peptide with the matrix used for MALDI analysis but may also represent the presence of a second, minor *O*-GlcNAc site.

To obtain further indication that the additional molecular mass of SpE expressed in vaccinia virus results from the addition of *O*-GlcNAc, we made use of ESMS. During ESMS analysis, ionization resulted in the release of a considerable fraction of the *O*-GlcNAc from the glycosylated peptides. Shown in Fig. 5B are the ESMS scans from 190 to 400 *m/z* of these peptides. While a signal corresponding to the released GlcNAc (204 mass units) was detected with SpE, no GlcNAc signal was detected with the SpS. Thus, the 203-mass-unit increase in the vaccinia virus SpE can be accounted for by a single GlcNAc. Moreover, these results further support the identity of the serine corresponding to Sp1 residue 484 as the dominant site of *O*-GlcNAcylation of the SpE peptide, because mutation of this site almost abolished glycosylation of the peptide.

**In vitro TAF110 and Sp1 protein association assays.** The Sp1 B-c domain has been shown to be involved in the homomultimerization of Sp1 (24) and in the interaction of Sp1 with TAF110 (8). By varying the state of SpE *O*-GlcNAcylation, we could study the effect of this modification on the ability of the SpE peptide to associate with either holo-Sp1 or TAF110. GST-SpE expressed in *E. coli* is not glycosylated nor is GST-SpS when it is expressed in vaccinia virus, whereas at least 70% of the GST-SpE molecules expressed in vaccinia virus are *O*-GlcNAcylated. Equal quantities of these fusion proteins were immobilized on glutathione-Sepharose beads and incubated with *Drosophila* TAF110 or human Sp1 that had been expressed as $^{35}$S-labeled proteins with a Promega TNT wheat germ extract kit. Wheat germ lysate does not *O*-GlcNAcylate proteins (31) and did not glycosylate the SpE expressed in *E. coli*. On the other hand, rabbit reticulocyte lysate has been shown to contain an *O*-GlcNAc transferase (31) and did glycosylate SpE to an undetermined extent. This difference in synthetic extracts was considered, because the interaction assay required that GST-SpE be exposed to these extracts during the binding studies, which possibly resulted in the glycosylation of the GST-peptide to an indeterminate extent. The use of the wheat germ system obviated this concern. The $^{35}$S-labeled holo-Sp1 or TAF110 contained in the TNT reaction was incubated with the immobilized GST-peptides, allowing us to directly compare the interactions with the variously *O*-GlcNAcylated forms of SpE. As shown in Fig. 6A, essentially equal quantities of the GST-fusion proteins were loaded onto the glutathione beads. GST alone and the glycosylated form of GST-SpE that had been expressed with vaccinia virus failed to bind either holo-Sp1 (Fig. 6B) or TAF110 (Fig. 6C) significantly. In contrast, the bacterially expressed unglycosylated GST-SpE, whose amino acid sequence was identical to that of the virally expressed GST-SpE, bound both Sp1 (Fig. 6B) and TAF110 (Fig. 6C). Similarly, the unglycosylated mutant GST-SpS, even though it was expressed in the vaccinia virus system, bound Sp1 (Fig. 6B) to approximately the same extent as the wild-type bacterially expressed peptide. Virally expressed GST-SpS also bound TAF110 (Fig. 6C) to a significantly greater extent than did virally expressed GST-SpE; however, SpS was less effective at binding TAF110 than unglycosylated SpE was. These results indicate that the mutation at the glycosylation site, while blocking glycosylation, was largely permissive for protein interactions with both holo-Sp1 and TAF110. That the unglycosylated wild-type and mutant model peptides still interacted with Sp1 and TAF110 suggests that the mutation at the glycosylation site did not significantly alter the conformation of the protein compared to that of the wild-type, unglycosylated form of SpE. These results indicate that O glycosylation of the Sp1 model peptide corresponding to the B-c activation domain prevents this glutamine-rich domain from hydrophobic interaction with the glutamine-rich interaction domains in TAF110 and Sp1.

**In vivo studies of the effect of the *O*-GlcNAc site mutation.** To determine if the mutation in SpS had an effect on transcription in an intact cell, both SpE and SpS were fused to the DNA-binding domain of Gal4 (amino acids 1 to 94). The fusion proteins were transiently expressed in HeLa cells under the control of the simian virus 40 promoter. The plasmids carrying genes encoding these fusion proteins were cotransfected into the cells with a luciferase reporter gene containing five copies of the Gal4 binding site placed upstream of a minimal promoter (Fig. 7). Unfused Gal4 activated this reporter minimally, while as reported before (8), Gal4-SpE was sixfold more potent. Gal4-SpS activated reporter function in a manner that was indistinguishable from that of SpE (Fig. 7). Gal4-SpE (8) and Gal4-SpS also activated the Gal4-dependent reporter similarly in *Drosophila* Schneider cells (data not shown), a cell line that does not express endogenous Sp1. Gal4-SpE, expressed with the vaccinia virus system, is glycosylated, as was determined by the galactosyltransferase method (data not shown), so we presume that the Gal4-SpE transiently expressed in HeLa cells is also *O*-GlcNAcylated.

## DISCUSSION

The domain structure of the transcription factor Sp1 has been well-defined (3, 4, 20, 24). Two transcriptional activation

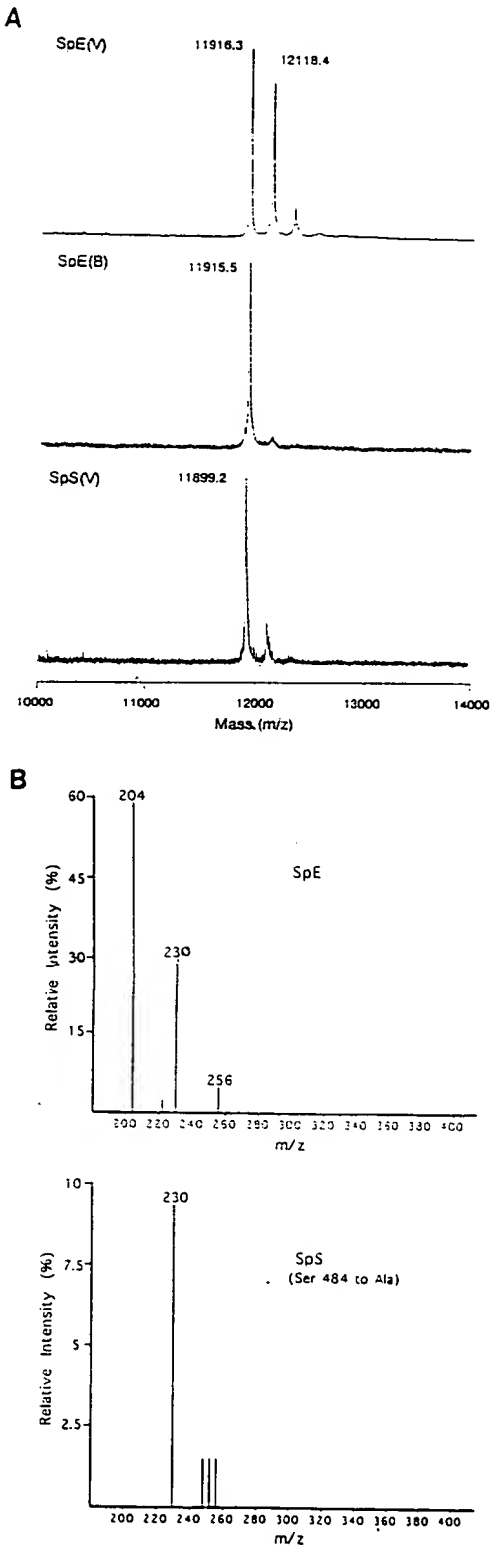


FIG. 5. Mass spectroscopic analysis of the SpE and SpS peptides. (A) MALDI-TOF analysis of the SpE and SpS peptides. The peptides were expressed as GST fusion proteins either in the vaccinia virus system (V) or in *E. coli* (B). The GST fusion proteins were affinity purified on glutathione-Sepharose and then cleaved from GST with thrombin. The released peptides were analyzed on a MALDI-TOF mass spectrometer. Indicated on the mass spectra are the molecular masses of these proteins as determined against external molecular weight standards. The predicted average molecular mass of unmodified SpE is 11,919.4 Da, that of unmodified SpS is 11,903.4 Da, and that of SpE modified with a single *O*-GlcNAc is 12,122.4 Da. (B) ESMS analysis of SpE and SpS. Wild-type SpE and the peptide containing a serine 484-to-alanine mutation (SpS) were expressed as GST fusion proteins in the vaccinia virus system. The peptides were cleaved from the GST and HPLC purified prior to ESMS. The upper graph shows the mass spectrum through the 190- to 400-*m/z* range of SpE and indicates the presence of the 204-*m/z* GlcNAc signal derived from SpE. The lower graph is a scan of the SpS peptide at the same range and indicates the absence of the 204-Da GlcNAc signal.

domains have been mapped to the N-terminal region of the molecule. These domains, termed the A and B domains, are glutamine rich and flanked C terminally by serine-threonine-rich regions. The B domain has been shown to confer two properties on Sp1. First, the B domain is involved in the homomultimerization of Sp1, thereby allowing superactivation of Sp1-dependent transcription (24). Second, the B domain interacts with TFIID through a protein whose homolog in *Drosophila melanogaster* is TAF110 (8, 15). These interactions appear to require hydrophobic residues in the glutamine-rich region of the B domain (8). These well-defined structure-function relationships in Sp1 prompted us to create a model system to test the role of O glycosylation in these transcriptionally relevant protein-protein interactions. The need for this model was dictated by this rather complex structure of Sp1. Sp1 is estimated to contain at least nine *O*-GlcNAc modification sites (19) and is also phosphorylated. These modifications may subserve different functions of Sp1. We therefore confined our studies to a well-circumscribed segment of Sp1, the B activation domain. When the B domain of Sp1 (SpE) was expressed in primate cells with recombinant vaccinia virus, we found that it was predominantly modified by a single *O*-GlcNAc residue at a serine corresponding to serine 484 in holo-Sp1. This site was established both directly by sequencing the [$^3$H]galactose-labeled peptide fragment that had been identified by mass spectroscopy and indirectly by mutating the modified serine residue and showing that the SpE peptide could no longer be glycosylated efficiently. This site of Sp1 *O*-GlcNAcylation was close to a homologous site of glycosylation that had been mapped in the activation domain of the SRF (26), compatible with a similar role for *O*-GlcNAcylation in Sp1 and SRF transcriptional activation. While *O*-GlcNAc sites have been mapped in other proteins that had been expressed in similar systems (1, 26), we cannot definitively state that the site we mapped in this peptide corresponds exactly to the site in the native protein. Nevertheless, the site in SpE appears to have been glycosylated with considerable specificity in that neither GST alone nor a similar-size segment of Sp1 was glycosylated. SpS was inefficiently glycosylated, and SpE was glycosylated in this system when it was expressed as a fusion partner with the DNA-binding domain of Gal4. In any event, this mapping allowed us to characterize and mutate the SpE peptide so that we could use it as a model to determine the effect of *O*-GlcNAc on SpE's known protein interactions.

The observation that the B domain of Sp1 is *O*-GlcNAcylated when it is expressed with vaccinia virus provided us the opportunity to directly assess the effect of this *O*-GlcNAcylation in Sp1-protein interactions. The interactions of SpE with either holo-Sp1 or TAF110 were assessed by an in vitro protein interaction assay. Similar assay systems have been used to assess transcriptionally relevant interactions between other proteins (10, 27), and results have been confirmed by in vivo studies (28). The in vitro interaction conditions were optimized to prevent nonspecific interactions between TAF110 or Sp1 and the glutathione-Sepharose beads or unfused GST protein.

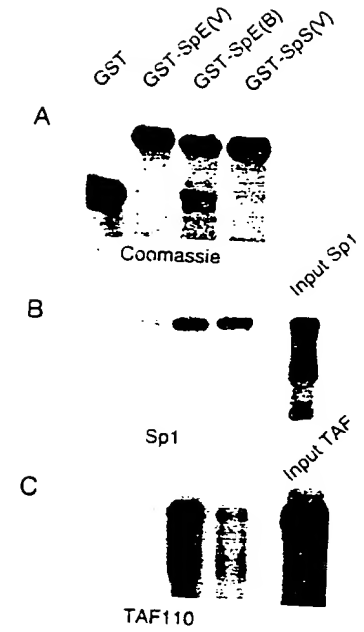


FIG. 6. In vitro protein interaction study between Sp1 or *Drosophila* TAF110 and the SpE peptides. SpE was expressed as a GST fusion protein with either vaccinia virus (V) or *E. coli* (B). The vaccinia virus-expressed GST-SpE was O glycosylated to a stoichiometry of at least 70% at serine 484, while the bacterially expressed peptide contained no *O*-GlcNAc. GST-SpS (serine 484 to alanine) was expressed in vaccinia virus and was not detectably glycosylated. The indicated GST peptides were bound in equal quantities to glutathione-Sepharose beads (A) and incubated with either $^{35}$S-Sp1 (B) or $^{35}$S-TAF110 (C) that had been expressed in a wheat germ extract system. The signals from 25% of the labeled input proteins are indicated in the last lanes of panels B and C. After extensive washing of the beads, the bound proteins were separated by SDS-PAGE. The GST peptides were observed following Coomassie blue staining of the gel, while the $^{35}$S-labeled proteins were observed by fluorography.

Under these conditions, the interaction studies indicated a distinct difference between the behavior of *O*-GlcNAcylated SpE and that of its unmodified form. Whether the protein was unmodified as a result of its expression in *E. coli* or as a result of a point mutation at the glycosylation site, the unmodified state was permissive for the interactions of the B domain with holo-Sp1 and TAF110. In sharp contrast, when the model peptide was *O*-GlcNAcylated, binding of both holo-Sp1 and TAF110 was virtually completely abolished. This protein interaction model based on the B domain of Sp1 strongly suggests that one role for Sp1 *O*-GlcNAcylation is in the control of protein-protein interactions.

The in vitro interaction studies indicate that a mutation of the predominant *O*-GlcNAc site in SpE is permissive for interaction with holo-Sp1 and TAF110. In HeLa cells, the mutation at this site in SpE did not have a significant impact on the activation of transcription, as was shown by the ability of the SpS peptide, when fused to Gal4, to activate a Gal4-dependent reporter. However, the in vitro results predict that *O*-GlcNAcylated SpE should not activate transcription because the peptide, in this form, does not interact efficiently with Sp1 or TAF110. Yet, we and others (8) have shown that this modifiable segment of Sp1 does activate transcription both in HeLa and in *Drosophila* cells. Broadly, there are two explanations for the indistinguishable behaviors of SpE and SpS in these cells. First, the glycosylation states of SpE and SpS in vivo may be the same, because either SpE is never glycosylated, SpE is deglycosylated, or SpS is glycosylated at some other site. Second, the in vivo glycosylation states of SpE and SpS may indeed differ, but the interactions required for transcriptional activation are not sensitive to the glycosylation state. For the in vivo system, we cannot distinguish these possibilities. However, we have shown that Sp1 does become hypoglycosylated when it is exposed to nuclear extract (12), and an *O*-GlcNAc-specific glucosaminidase in the nucleus has been described (6). Thus, the *O*-GlcNAc modification of Gal4-SpE may be transient in vivo (see below), thereby permitting transcriptional activation by the wild-type peptide.

The protein-protein interactions that are required for the creation of a transcriptionally competent complex at the appropriate position on the DNA can result in a very stable complex. With the TFIID complex, which includes the TATA-binding protein and associated TAFs, these protein interactions are sufficiently stable in an aqueous environment to prevent their dissociation except by the use of chaotropic agents (32). The intrinsic stability of these protein complexes raises an interesting problem with regard to those nuclear proteins like Sp1 that homomultimerize. These proteins are synthesized on polysomes in the cytoplasm and thereby attain significant local concentrations during and immediately after synthesis. The problem is, then, how these proteins are prevented from assembling into homomultimeric complexes as they are presumed to do in the nucleus upon binding DNA. In this paper, we demonstrate that *O*-GlcNAcylation of the Sp1 B domain model peptide prevents the interaction of this peptide with its parent protein, Sp1. Removal of the sugar either by mutagenesis or by the expression of the peptide in a system that does not *O*-GlcNAcylate is permissive for protein interaction. Assuming that the behavior of the Sp1 model peptide may be representative of that of the holoprotein, then O-GlcNAcylation may be a means of preventing inappropriate homomultimerization during or after translation in the cytoplasm. Since *O*-GlcNAcylation has been shown to occur cotranslationally, at

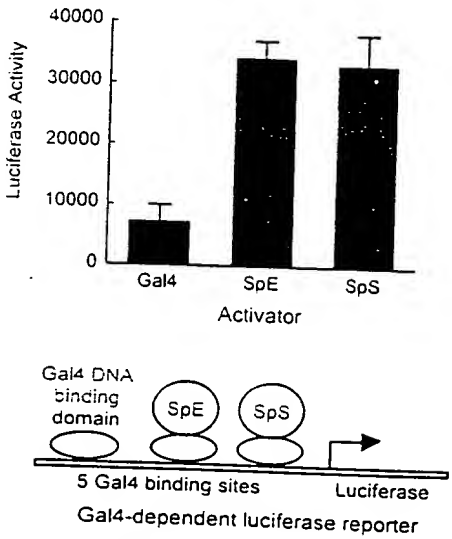


FIG. 7. Effect on transcription in HeLa cells of an *O*-GlcNAc site mutation in the SpE peptide. HeLa cells were transiently transfected with a Gal4-dependent reporter plasmid along with expression plasmids carrying genes encoding the Gal4 DNA-binding domain fused or not fused to the cDNA encoding the indicated SpE peptide. The cells were also cotransfected with a cytomegalovirus β-galactosidase reporter plasmid to control for transfection efficiency. Luciferase activities were determined in cell extracts 24 h following the transfection. All transfections were performed in triplicate. The diagram below the bar graph indicates a schematic representation of the reporter system used in this experiment.

least for the nucleoporin p62 (31), then the problem of Sp1 multimerization may be averted by the *O*-GlcNAcylation of this protein during its synthesis on the polysome. The prevention of cytoplasmic complex formation through *O*-GlcNAcylation may be necessary for proteins destined for the nucleus, since the nuclear pore does restrict access to the nucleus.

Within the nucleus, problems may also arise as a result of untimely or spatially inappropriate protein complex formation. Following entry of Sp1 into the nucleus, Sp1 may interact in the nucleosol with the same proteins with which it normally interacts on the DNA template. There is indeed some indirect evidence that such inappropriate interactions can occur under experimental conditions. The phenomenon of squelching, in which general transcription, even from genes lacking Gal4-binding sites, is impaired by the overexpression of Gal4, was described some years ago (9). This squelching phenomenon was believed to result from the interaction of the Gal4 activation domain with soluble general transcription factors, thereby titrating these factors away from the genes with which they normally interact to activate transcription. Extrapolating from our Sp1 peptide model, we propose that *O*-GlcNAcylation of Sp1 and perhaps other transcription factors might prevent such ectopic protein-protein interactions. The corollary of this proposal requires that the *O*-GlcNAc modification must be removed to allow the essential interactions between transcription factors to occur on the DNA template. We speculate that the removal of *O*-GlcNAc from the interaction domain may occur in a DNA-dependent manner upon the binding of Sp1 to the GC box. Since Sp1 is already known to undergo phosphorylation upon its binding to DNA (18), other DNA-dependent modifications of the protein may occur in conjunction with phosphorylation. Indeed, there is indirect evidence supporting the concept that *O*-GlcNAcylation may flip-flop with phosphorylation. Examples are the *O*-GlcNAc site in c-Myc that corresponds to a known site of phosphorylation (1) and the carboxy-terminal domain of RNA polymerase II that appears to alternate between phosphorylation and *O*-GlcNAcylation (22). Our model would predict the existence of a hexosaminidase whose activity towards Sp1 is controlled in a manner that senses the binding of Sp1 to the GC box on the DNA. The removal of the sugar would then permit DNA-dependent transcriptionally relevant protein complex formation.

This model for the role of *O*-GlcNAc may also contribute to the specificities of protein interactions. If it were necessary for Sp1 to bind to DNA before it was deglycosylated, then the *O*-GlcNAc on the Sp1 not bound to DNA would prevent Sp1 from entering into nonspecific hydrophobic interactions with other transcription factors that do not bind to the same promoter but contain glutamine-rich activation domains. While other structural determinants in Sp1 may also contribute to this specificity, the occlusion of the interaction domain by *O*-GlcNAc until Sp1 bound DNA would increase this specific-

We have recently proposed another role for Sp1 *O*-GlcNAcylation. We created, through glucose deprivation and cAMP activation, a condition in cells that resulted in the hypoglycosylation of Sp1 and other nuclear proteins. Under these conditions of hypoglycosylation, we observed the rapid proteolysis of Sp1 by a proteasome-like mechanism (12). We interpreted this degradation of Sp1 as a means of shutting down general transcription of housekeeping genes under stress conditions to conserve nutrients. Whether the control of this proteolytic process is related to the role of *O*-GlcNAc in preventing protein complex formation is not clear. However, it remains possible, based on the Sp1 model described in this paper, that hypoglycosylated Sp1 enters into protein complexes that are recognized by the proteasome as inappropriate (2). Taken together, glycosylation may allow Sp1 to remain stable and free of complex formation. This soluble form of Sp1 may serve as a latent reservoir of the transcription factor. Such a reservoir might be required at points in the cell cycle at which transcription is quiescent, such as during the S phase and mitosis. The cell cycle-dependent (13) and signal-dependent (21) changes that have been observed in protein *O*-GlcNAcylation might reflect this postulated role for this protein modification. Although our studies provide direct evidence for a role of *O*-GlcNAcylation in prevention of protein interaction, there are other possible roles for *O*-GlcNAcylation that have been hypothesized (13). Indeed, because Sp1 is multiply glycosylated, there may be other roles for this modification even within this one molecule.

## ACKNOWLEDGMENTS

We thank Gerald Hart, Shane Arnold, and Richard Marchase for their help and advice with this study.

This work was supported by a grant from the Juvenile Diabetes Foundation.

## REFERENCES

1. **Chou, T. Y., G. W. Hart, and C. V. Dang.** 1995. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. J. Biol. Chem. 270:18961–18965.
2. **Ciechanover, A.** 1994. The ubiquitin-proteasome proteolytic pathway. Cell 79:13–21.
3. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell 59:827–836.
4. **Courey, A. J., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55:887–898.
5. **Cress, W. D., and S. J. Triezenburg.** 1991. Critical structural elements of the VP16 transcriptional activation domain. Science 251:670–682.
6. **Dong, D. L.-Y., and G. W. Hart.** 1994. Purification and characterization of an *O*-GlcNAc selective N-acetyl-β-D-glucosaminidase from rat spleen cytosol. J. Biol. Chem. 269:19321–19330.
7. **Ferreri, K., G. Gill, and M. Montminy.** 1994. The cAMP-regulated transcription factor CREB interacts with a component of the TFIID complex. Proc. Natl. Acad. Sci. USA 9:1210–1213.
8. **Gill, G., E. Pascal, Z. H. Tseng, and R. Tjian.** 1994. A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. Proc. Natl. Acad. Sci. USA 91:192–196.
9. **Gill, G., and M. Ptashne.** 1988. Negative effect of the transcriptional activator Gal4. Nature (London) 334:721–724.
10. **Goodrich, J. A., T. Hoey, C. J. Thut, A. Admon, and R. Tjian.** 1993. TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. Cell 75:519–530.
11. **Haltiwanger, R. S., M. A. Blomberg, and G. W. Hart.** 1992. Glycosylation of nuclear and cytoplasmic proteins. J. Biol. Chem. 267:9005–9013.
12. **Han, I., and J. E. Kudlow.** 1997. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. Mol. Cell. Biol. 17:2550–2558.
13. **Hart, G. W.** 1997. Dynamic *O*-GlcNAcylation of nuclear and cytoskeletal proteins. Annu. Rev. Biochem. 66:315–335.
14. **Hart, G. W., R. S. Haltiwanger, G. D. Holt, and W. G. Kelly.** 1989. Glycosylation in the nucleus and cytoplasm. Annu. Rev. Biochem. 58:841–871.
15. **Hoey, T., R. O. J. Weinzierl, G. Gill, J. L. Chen, B. D. Dynlacht, and R. Tjian.** 1993. Molecular cloning and functional analysis of *Drosophila* TAF110 reveals properties expected of coactivators. Cell 72:247–260.
16. **Holt, G. D., and G. W. Hart.** 1986. The subcellular distribution of terminal *N*-acetylglucosamine moieties. J. Biol. Chem. 261:8049–8057.
17. **Holt, G. D., C. M. Snow, A. Senior, R. S. Haltiwanger, L. Gerace, and G. W. Hart.** 1987. Nuclear pore complex glycoproteins contain cytoplasmically disposed *O*-linked *N*-acetylglucosamine. J. Cell Biol. 104:1157–1164.
18. **Jackson, S. P., J. J. MacDonald, S. Lees-Miller, and R. Tjian.** 1990. GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. Cell 63:155–165.
19. **Jackson, S. P., and R. Tjian.** 1988. *O*-Glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell 55:125–133.
20. **Kadonaga, J. T., A. J. Courey, J. Ladika, and R. Tjian.** 1988. Distinct regions of Sp1 modulate DNA binding and transcriptional activation. Science 242: 1566–1570.

21. **Kearse, K. P., and G. W. Hart.** 1991. Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins. Proc. Natl. Acad. Sci. USA 88:1701–1705.
22. **Kelly, W. G., M. E. Dahmus, and G. W. Hart.** 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by *O*-GlcNAc. J. Biol. Chem. 268:10416–10424.
23. **Moss, B.** 1991. Vaccinia virus: a tool for research and vaccine development. Science 252:1662–1667.
24. **Pascal, E., and R. Tjian.** 1991. Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes Dev. 5:1646–1656.
25. **Pugh, B. F., and R. Tjian.** 1991. Transcription from a TATA-less promoter requires a multi-subunit TFIID complex. Genes Dev. 5:1935–1945.
26. **Reason, A. J., H. R. Morris, M. Panico, R. Marais, R. H. Treisman, R. S. Haltiwanger, G. W. Hart, W. G. Kelly, and A. Dell.** 1992. Localization of *O*-GlcNAc modification on the serum response transcription factor. J. Biol. Chem. 267:16911–16921.
27. **Sauer, F., S. K. Hansen, and R. Tjian.** 1995. DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* Bicoid. Science 270:1825–1828.
28. **Sauer, F., D. A. Wassarman, G. M. Rubin, and R. Tjian.** 1996. TAFs mediate activation of transcription in the *Drosophila* embryo. Cell 87:1271–1284.
29. **Schindler, M., M. Hogan, R. Miller, and D. DeGaetano.** 1987. A nuclear specific glycoprotein representative of a unique pattern of glycosylation. J. Biol. Chem. 262:1254–1260.
30. **Shin, T. H., and J. E. Kudlow.** 1994. Identification and characterization of the human transforming growth factor α initiator. Mol. Endocrinol. 8:704–712.
31. **Starr, C. M., and J. A. Hanover.** 1990. Glycosylation of nuclear pore protein p62. Reticulocyte lysate catalyzes *O*-linked *N*-acetylglucosamine addition in vitro. J. Biol. Chem. 265:6868–6873.
32. **Tanese, N., B. F. Pugh, and R. Tjian.** 1991. Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. 5:2212–2224.
33. **Torres, C., and G. W. Hart.** 1984. Topography and polypeptide distribution of terminal *N*-acetylglucosamine residues on the surface of intact lymphocytes. J. Biol. Chem. 259:3308–3317.

ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS
Vol. 325, No. 1, January 1, pp. 8–19, 1996
Article No. 0002



# Expression of Biologically Active Human SPARC in *Escherichia coli*[1]

James A. Bassuk,[2] François Baneyx,* Robert B. Vernon, Sarah E. Funk, and E. Helene Sage
*Departments of Biological Structure and *Chemical Engineering, University of Washington, Seattle, Washington 98195*

Received August 8, 1995, and in revised form October 16, 1995

**Human SPARC has been cloned by the polymerase chain reaction from an endothelial cell cDNA library and expressed in *Escherichia coli* as a biologically active protein. Transcriptional expression of the insert cDNA was dependent on the activation of the T7 RNA polymerase promoter by isopropylgalactopyranoside. Two forms of recombinant SPARC (rSPARC) protein were recovered from BL21(DE3) *E. coli* after transformation with the plasmid pSPARCwt: a soluble, monomeric form of rSPARC and an insoluble, aggregated form sequestered in inclusion bodies. The isolation of the soluble form of rSPARC was accomplished by anion-exchange, nickel-chelate affinity, and gel filtration chromatographies. The isolated protein was an intact, full-length polypeptide of 293 amino acids by the following criteria: N-terminal amino acid sequence, reaction with anti-SPARC immunoglobulins specific for N-terminal and C-terminal sequences, and interaction of the C-terminal histidine tag of rSPARC with a nickel-chelate affinity resin. Circular dichroism and intrinsic fluorescence spectroscopy indicated that the conformation of rSPARC was dependent on interaction with $Ca^{2+}$ ions. The recombinant protein inhibited cell spreading and bound specifically to bovine aortic endothelial cells. Levels of bacterial endotoxin (<18 pg/μg rSPARC) present in rSPARC preparations were below the threshold that affects the behavior of these endothelial cells. These conformational and biological properties of rSPARC are consistent with previously described characteristics of the native protein. The purification of biologically active rSPARC, as well as mutated forms of the protein, will provide sufficient quantities of protein for the determination of structure/function relationships.** © 1996 Academic Press, Inc.

***Key Words:*** **SPARC; BM-40; osteonectin; recombinant protein expression; bacteria; protein folding; nickel-chelate affinity chromatography; vascular biology; extracellular matrix.**

SPARC (secreted protein acidic and rich in cysteine) is a secreted glycoprotein that interacts with cells and extracellular matrix components (reviewed in 1, 2). The expression pattern of SPARC during vertebrate embryogenesis, as well as in tissues undergoing remodeling and repair, is consistent with a fundamental role for this protein in tissue morphogenesis and cellular differentiation (3). SPARC, the translation product of a single gene (4), has also been termed BM-40 (5) and osteonectin (6). To date, the precise function of SPARC has remained elusive. A recent report has indicated that SPARC is a source of copper-binding peptides that regulate angiogenesis (7). The specific degradation of SPARC into fragments of diverse biological activity has provided an attractive model by which SPARC might affect vascular growth and remodeling.

Preparations of SPARC from cells and tissues have often been compromised by the activities of endogenous proteinases (5). In addition, SPARC is conformationally labile and undergoes significant loss of activity, presumably because denaturing conditions are often employed for the extraction and purification of the protein (5, 8, 9). This instability is likely due to flexible segments of the protein (10). Moreover, SPARC isolated from human platelets (11), or from the conditioned media of cells (9), is often contaminated with platelet-derived growth factor, a mitogen and chemokine that binds specifically to SPARC (12).

Together with its acidic p*I* of 4.5 and high content of nonpolar amino acids (41%), several characteristics of this intriguing polypeptide indicate a dynamic structure that responds to fluctuations in physiochemical

[1] This work was supported by a Grant-in-Aid from the American Heart Association Washington Affiliate (94-WA-501) and National Institutes of Health Grants P50-DK-47659, GM-40711, and HL-18645.

[2] To whom correspondence should be addressed at the Department of Biological Structure, University of Washington School of Medicine, Box 357420, Seattle, WA 98195-7420. Fax: (206) 543-1524; E-mail: Bassuk@u.washington.edu.



0003-9861/96 $12.00

stimuli. To understand how the various structural motifs of SPARC determine its functional properties, we have designed a bacterial expression system to acquire quantities of the recombinant protein sufficient for biochemical studies. We report here that our preparations of SPARC display both biological and conformational activities that are consistent with those of the native protein produced by mammalian cells.

## EXPERIMENTAL PROCEDURES

*Materials.* A human umbilical vein endothelial cell cDNA library was provided by Cathy Lofton-Day of Zymogenetics, Inc. (Seattle, WA). The plasmid pET22b and *Escherichia coli* strain BL21(DE3) were purchased from Novagen (Madison, WI). Fetal bovine serum (FBS)[3] was purchased from HyClone (Logan, UT) and Dulbecco's modified Eagle's medium (DMEM) was obtained from Sigma (St. Louis, MO). Restriction endonucleases *Nde*I and *Xho*I, T4 DNA ligase, and *Taq* polymerase were purchased from Boehringer Mannheim (Indianapolis, IN). Ni–NTA metal-chelate affinity resin was purchased from Qiagen (Chatsworth, CA) and DEAE-Sepharose Fast Flow ion-exchange resin and Superdex-70 gel filtration resin from Pharmacia (Piscataway, NJ). Disposable gel filtration columns (Econopak 10DG), Chelex-100, and precast SDS–PAGE gels were acquired from Bio-Rad (Foster City, CA). Ampicillin, penicillin G, streptomycin sulfate, and carbenicillin were obtained from Sigma, and [$^{35}$S]dATP, $^{125}$I-protein A, and [$^{125}$I]NaI from Amersham (Arlington Heights, IL). AEBSF was purchased from Calbiochem (La Jolla, CA). A *Limulus* amebocyte lysate QCL-1000 kit was obtained from BioWhittaker (Walkersville, MD), and a BCA protein assay and Iodo-Bead kit were purchased from Pierce (Rockford, IL).

*Construction of recombinant plasmid encoding human endothelial SPARC.* Total DNA was isolated from a human umbilical vein endothelial cell pZEM-228cc cDNA library and was used as a template for *Taq* polymerase in a polymerase chain reaction (PCR). One pair of oligonucleotide primers was synthesized: L, 5′-CATCCACGGCATATGGCCCCTCAGCAAGAAG-3′, and R, 5′-ACTCTCCGGCTACTCGAGGATCACAAGATCCTTG-3′. One unit of *Taq* polymerase was added and 2 cycles (94°C, 1.5 min; 55°C, 1 min; and 72°C, 3 min), 23 cycles (94°C, 1 min; 72°C, 2.5 min), and 1 cycle (74°C, 7 min) of PCR were performed. The resultant PCR-DNA product was resolved by electrophoresis through an agarose gel, and a band of 900 bp was excised and isolated with β-agarase. The oligonucleotide primer L conferred the recognition site for the restriction endonuclease *Nde*I 5′ to the first translation codon (Ala) of the rSPARC PCR-DNA product. Oligonucleotide primer R conferred the site for *Xho*I at the 3′ end of the rSPARC PCR-DNA product. The PCR-DNA product and the bacterial expression plasmid pET-22b were each digested with *Nde*I and *Xho*I and were joined with T4 ligase to form the recombinant plasmid pSPARCwildtype (pSPARCwt). Competent BL21(DE3) *E. coli* were transformed with pSPARCwt, and stable cultures were propagated and frozen as glycerol stocks. A large-scale preparation of pSPARCwt was achieved by growth in 0.5 liter liquid culture, alkaline lysis of the bacteria, and isolation of closed, circular pSPARCwt by centrifugation through a gradient of CsCl. Dideoxy DNA sequencing and subse analysis confirmed that the PCR-DNA product encoded h SPARC and that the coding region was placed in the correct re frame for proper expression. Human SPARC/BM-40/osteon mRNA sequences used for comparison with pSPARCwt were viously reported (GenBank P09486) (13, 14).

*Growth and expression of recombinant SPARC cultures.* An night culture of pSPARCwt in BL21(DE3) *E. coli* was inoculat a dilution of 1:50 into Luria broth (1% w/v bacto-tryptone, 0.5% bacto-yeast extract, 0.5% w/v NaCl) supplemented with 0.2% glucose and 50 μg/ml carbenicillin and was grown at 37°C in a Brunswick BioFlo-C32 fermentor with full aeration and agitatio at 500 rpm. The pH of the medium was maintained at 7.0 by a matic addition of HCl or NaOH. The working volume was 1.3 li Isopropylgalactopyranoside (IPTG) was added to a final concen tion of 1 mM to midexponential phase cells ($OD_{600}$ ~ 0.5), and culture was grown for an additional 3.5 h. The cells were recov by centrifugation at 7000*g*, resuspended in 20 ml of 10 mM sod phosphate (pH 7.0) that contained 10% glycerol, and disrupted two cycles in a french press (20,000 psi). Soluble and insoluble m rials were separated by centrifugation at 10,000*g* for 30 min at 4 Soluble extracts and pellets were stored at −80°C.

*Isolation of rSPARC protein.* Soluble extracts were thawed on and diluted to 100 ml with buffer A [90 mM sodium phosphate ( 7.8), 10% glycerol (v/v)] that contained 0.2 mM AEBSF, a wat soluble inhibitor of serine proteases. Fifty milliliters of settled DEA Sepharose Fast Flow anion-exchange resin, previously equilibrat in the same buffer, was added and the mixture was stirred gen for 12–18 h at 4°C. The slurry was subsequently poured into a ch matography column (2 × 20 cm), allowed to settle at 1*g*, and wash with approximately 250 ml of buffer A until the $OD_{280}$ was <0.0 The column was developed at 3 ml/min with a linear gradient of 2 ml of buffer A to 250 ml of buffer A that contained 0.5 M NaC Fractions of 8 ml were collected. The column eluate was monitore with a flow cell coupled to a uv monitor at 280 nm and to a cha recorder (full scale = 1 OD). The ionic strength of the eluate wa monitored with a conductivity meter. Under these condition rSPARC eluted at a salt concentration of 0.10–0.25 M NaCl (conduc tivity of 14–20 mmho). These fractions were analyzed rapidly b SDS–PAGE of 50-μl aliquots (minigels of 5 × 5 cm), prior to storag at −80°C. Fractions that contained rSPARC were pooled and wer adjusted to 0.5 M NaCl by the addition of 5 M NaCl until the conduc tivity of the sample was equivalent to the conductivity of buffer B (0.5 M NaCl, 50 mM sodium phosphate, 10% glycerol, pH 7.8). Alternatively, bound proteins were eluted from the DEAE-Sepharose resin with 50 ml of buffer A that contained 0.5 M NaCl. AEBSF was added to a final concentration of 0.2 mM from a stock solution of 0.2 M. The sample was added to 3–5 ml (per liter of original culture) of a 50% slurry of Ni–NTA metal-chelate affinity resin (Qiagen), and the mixture was adjusted to pH 7.8. After gentle stirring for 1 h at 4°C, the slurry was poured into a chromatography column 1 × 10 cm), allowed to settle at 1*g*, and washed with approximately 60 ml of buffer B at 0.5 ml/min until the $OD_{280}$ was <0.01. Proteins that were nonspecifically bound to the resin or to rSPARC were removed with 15 column volumes of buffer B adjusted to pH 6.0 or until the $OD_{280}$ was <0.01. rSPARC was subsequently eluted from the column with 20 ml of buffer B adjusted to pH 5.3 and was stored in aliquots at −80°C.

Monomers of SPARC were separated from dimers, trimers, and oligomers by chromatography at 4°C through Superdex-70 gel filtration resin (1.6 × 60 cm) in 50 mM Tris–HCl (pH 8.0), containing 0.15 M NaCl, at a flow rate of 0.1 ml min$^{-1}$. Under these conditions, the column was calibrated for the elution volumes of blue dextran, cytochrome *c*, and [$^3$H]proline.

Buffer exchange was accomplished by the use of sterile, disposable, 10-ml gel filtration columns. Samples of rSPARC (0.1–0.5 mg/ml in Ni–NTA column elution buffer) were thawed, diluted into the new

[3] Abbreviations used: AEBSF, aminoethylbenzenesulfonyl fluoride; BAE, bovine aortic endothelial; BCA, 2,2′-bicinchoninic acid copper protein reagent; BSA, bovine serum albumin; CD, circular dichroism spectroscopy; DEAE, diethylaminoethyl; DMEM, Dulbecco's modified Eagle's medium; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FBS, fetal bovine serum; IDA, iminodiacetate; IPTG, isopropylgalactopyranoside; Ni–NTA, nickel–nitrilotriacetate; PCR, polymerase chain reaction; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

 

buffer to 3 ml, and applied to the top of the column. The exchange buffer, e.g., DMEM, was applied to the column (10 ml), and the protein eluting in the void volume was detected by a flow cell coupled to a uv monitor (280 nm) and a chart recorder. The use of sterilized exchange buffer resulted in preparations of rSPARC that were suitable for cell culture.

Preparations of rSPARC used in tissue culture experiments were evaluated for bacterial endotoxin content by the *Limulus* amebocyte lysate assay (15). Samples and standards were incubated with lysate for 10 min at 37°C in a microtiter plate and subsequently with Ac-Ile-Glu-Ala-Arg-*p*-nitroaniline for 6 min. The reaction was terminated by the addition of glacial acetic acid to a final concentration of 8.3% (v/v). The release of *p*-nitroaniline was measured photometrically by the absorbance at 405 nm.

The amount of rSPARC at each step of the purification was determined by densitometric scanning of SDS–polyacrylamide electrophoretic gels stained with Coomassie brilliant blue R-250. A standard curve was generated by plotting the percentage of area of rSPARC, calculated from the scanning of photographic negatives of 1, 5, and 10 $\mu$g rSPARC at 600 nm with a Beckman DU spectrophotometer, vs protein mass. Photographic negatives of SDS gels containing soluble lysates and DEAE-Sepharose fractions were used to calculate milligrams of rSPARC and the percentage of area for the 34- to 38-kDa fraction.

*Circular dichroism spectroscopy.* rSPARC (600 $\mu$g) and murine PYS-SPARC (600 $\mu$g) were dissolved into 2 ml of a Chelex-treated buffer consisting of 50 mM Tris–HCl (pH 8.0) and 150 mM NaCl. $CaCl_2$ was added to 50 mM, and $Ca^{2+}$ not bound to the proteins was removed by gel filtration chromatography. Proteins were diluted to 50 $\mu g\ ml^{-1}$ of the same buffer containing varying concentrations of EDTA, placed in quartz cells of pathlength 0.1 mm, and scanned repetitively (5–15×) from 290 to 195 nm in a Shiatsu J700 circular dichroism spectrometer at ambient temperature. The spectra of the Chelex-treated buffer were subtracted from the spectra of buffers that contained protein.

*Fluorescence emission spectroscopy.* Fluorescence measurements were performed in a Perkin–Elmer LS-50B luminescence spectrometer controlled by FLDM software operating on a DEC 386SX computer. rSPARC was dissolved at a concentration of 30 $\mu g\ ml^{-1}$ into a Chelex-treated 0.01 M Tris–HCl (pH 7.8) buffer that contained 0.15 M NaCl. A duplicate sample was prepared that contained 10 mM $CaCl_2$. Each sample was placed in a semimicrocuvette of pathlength 1 cm and was excited, in separate experiments, at ultraviolet light wavelengths of 260, 270, 275, 280, 285, 290, 300, and 310 nm. Emission spectra were monitored from 300 to 620 nm at a scan speed of 200 nm $min^{-1}$. Excitation slitwidths were 6 and 7 nm for excitation and emission, respectively. All measurements were performed at ambient temperature. The spectra of solutions that contained no protein were subtracted from the spectra of those containing protein. This subtraction removed the Raman spectra of water intrinsic to all the samples. To [illegible] to the observed emission spectra, we determined the fluorescent emission profiles of Trp, Tyr, and Phe at an excitation wavelength of 280 nm. Trp was identified as the residue that emitted fluorescence light typical of the spectra observed for the recombinant protein (16). First- and second-order Raleigh scattering were observed in all spectra and were separated optimally from the Trp signal at an excitation wavelength of 280 nm.

For titration experiments, 400–600 $\mu$g of rSPARC or mouse PYS-SPARC (9) was saturated with 0.05 M $CaCl_2$. $Ca^{2+}$ not bound to the protein was removed by gel filtration chromatography in Chelex-treated 0.05 M Tris–HCl/0.15 M NaCl at pH 8. Protein solutions were diluted to 20–30 $\mu$g/ml in the same buffer that contained varying concentrations of EDTA. The percentage of change in fluorescence intensity at 336 nm was calculated as previously described (16).

*Assay of cellular spreading.* Bovine aortic endothelial (BAE) cells (9) were maintained at 37°C/5% $CO_2$ in DMEM that contained the following supplements: 10% FBS, 100 units/ml penicillin G, and 100 units/ml streptomycin sulfate. Cultures were passaged with trypsin/EDTA and were used prior to passage 12.

Human rSPARC (100 $\mu$g/ml) was dialyzed against DMEM/1% FBS and was added to the wells of 24-well plastic dishes at varying concentrations and at constant volume. Confluent BAE cells were released with trypsin, and $10^4$ cells in suspension was added to each well. After 2 h, cells were photographed and were subsequently scored for spreading by a semiquantitative rounding index previously established for BAE cells (17). Cells were scored as (a) spread, flattened cells with diminished cellular refractility, (b) unspread, rounded cells projecting short processes in the initial stages of spreading, and (c) round, highly refractile cells with no apparent processes. The number of cells in each group was then converted into a "rounding index" by the formula $[(1 \times a) + (2 \times b) + (3 \times c)]/(a + b + c)$. An index of 1 thus represents a culture with only spread cells. A culture with increasing numbers of unspread and round cells would approach the maximum rounding index of 3. Values were calculated for two independent cultures and were graphed as the average ± SE. Similar results were obtained when the assay was repeated with a different preparation of rSPARC. A control sample contained the same volume fraction of Ni–NTA elution buffer (pH 5.3).

*Binding of rSPARC to endothelial cells.* Thirty-five micrograms of rSPARC was radioiodinated in 100 mM phosphate buffer that contained 1 mCi [$^{125}$I]NaI and Iodo-Beads according to the manufacturer's protocol. Cell surface binding experiments were performed as previously described (18) with the following modifications: BAE cells were grown to confluency in 24-well plates and were washed two times with 1 ml of ice-cold medium (DMEM that contained 25 mM Hepes, pH 7.4, and 2.5 mg/ml BSA). Varying amounts of $^{125}$I-rSPARC were added to a final volume of 0.5 ml binding buffer, with or without the indicated amount of unlabeled competitor (SPARC synthetic peptide 4.2, $NH_2$-TCDLDNDKYIALEEWAGCFG-COOH). This peptide had been shown to compete with native SPARC for binding to BAE cells (19). After an incubation of 2 h at 4°C, cultures were washed three times with cold binding buffer, solubilized with 1% Triton X-100 that contained 1 mg/ml BSA, and counted in a Beckman Gamma 4000 $\gamma$ counter. Specific binding of rSPARC was calculated from competition experiments as described (19).

## RESULTS

### *Expression of Recombinant Human Endothelial SPARC*

The cDNA coding region of rSPARC, isolated by PCR from a human umbilical vein endothelial cell cDNA library, was ligated into the plasmid pET22b to produce pSPARCwt (Fig. 1). Double-stranded, dideoxy sequencing of both strands of pSPARCwt showed it to be 100% identical with human bone osteonectin (13, 14). The N-terminus of rSPARC contains an extra amino acid (Met) compared to the secreted form, which creates a translation initiation codon. The C-terminus of rSPARC contains eight additional amino acids: Leu, Glu, and $(His)_6$. The Leu and Glu residues resulted from the ligation of rSPARC into the *Xho*I restriction site of pET22b. The six His residues were engineered to provide a high-affinity metal-chelation site for purification. pSPARCwt was subsequently transformed into competent BL21(DE3) *E. coli*.

Recombinant *E. coli* were grown in liquid culture, and the pattern of protein expression was assessed by

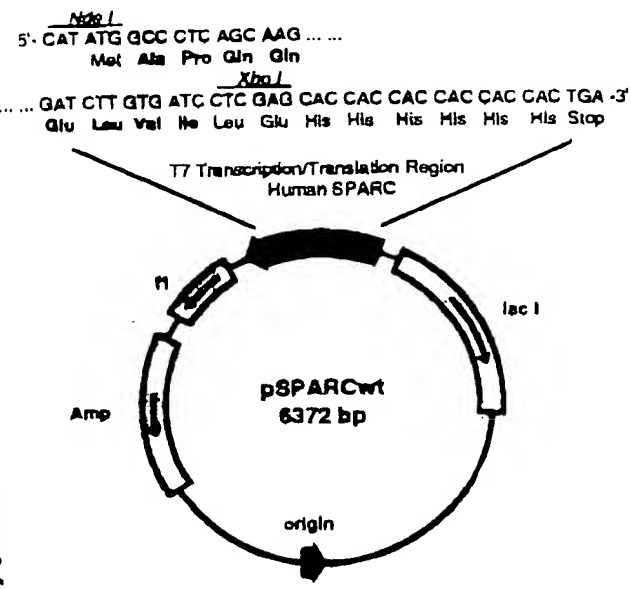


FIG. 1. The recombinant plasmid pSPARCwt. A 900-bp PCR-DNA product encoding human endothelial SPARC was inserted into the bacterial expression plasmid pET22b at restriction endonuclease sites *Nde*I and *Xho*I to produce the recombinant plasmid pSPARCwt. Shown are the genes for β-lactamase (Amp), the *lac* repressor (lacI), and single-strand replication (f1 origin). SPARC transcription is under the control of the T7 RNA polymerase promoter. Wild-type amino acids are designated in bold. Additional residues not found in the wild-type protein include an N-terminal Met and the C-terminal amino acids Leu, Glu, and $(His)_6$.

SDS–PAGE (Fig. 2). The expression of rSPARC was dependent on the addition of IPTG, which inactivates the *lac* repressor and induces the synthesis of chromosomally encoded T7 RNA polymerase under *lac*UV5 control and subsequent transcription of rSPARC mRNA from its T7 promoter. rSPARC appeared to be a prominent protein in induced cells after resolution of total bacterial protein by SDS–PAGE (Fig. 2). This electrophoretic profile was observed consistently whether rSPARC was expressed in shake-flask cultures or pilot-scale fermentations. The number of proteins present in fractions of equivalent volumes precluded our determining whether rSPARC was the major protein present in the IPTG-induced fractions. When total bacterial proteins were probed with immunoglobulins specific for the COOH-terminal EF-hand motif of rSPARC (Fig. 2B), an immunoreactive band was observed at the position that corresponded to the major IPTG-induced product in gels stained with Coomassie blue. We also used antibodies against an $NH_2$-terminal sequence of SPARC; the immunoblot appeared identical to the profile generated with COOH-terminal antibodies (data not shown). These results demonstrate that rSPARC of molecular mass 34–38 kDa was apparently expressed as a full-length polypeptide. Lane 6 of Fig. 2B also contains an immunoreactive band of ~97 kDa that might represent a form of rSPARC not fully reduced. rSPARC displays an electrophoretic mobility slightly greater than that of native SPARC isolated from mouse PYS-2 cells (Fig. 2B). This difference in apparent $M_r$ is due to an absence of carbohydrate in the bacterially expressed protein, whereas the protein expressed by mammalian cells is known to be glycosylated (11, 20).

### *Isolation and Characterization of Soluble rSPARC*

We were able to purify rSPARC from lysates of bacterial fermentations by sequential chromatography on ion-exchange and nickel-chelate resins. Figure 3 displays the electrophoretic profile of the purified protein. During the development of the purification protocol for rSPARC, we discovered that the manner in which we exposed the bacterial lysate to the ion-exchange resin was important and summarize our findings below.

*Resolution of rSPARC aggregates on ion-exchange resin.* Our initial attempts to isolate soluble rSPARC resulted in relatively poor yields. because varying amounts of the protein precipitated dur-

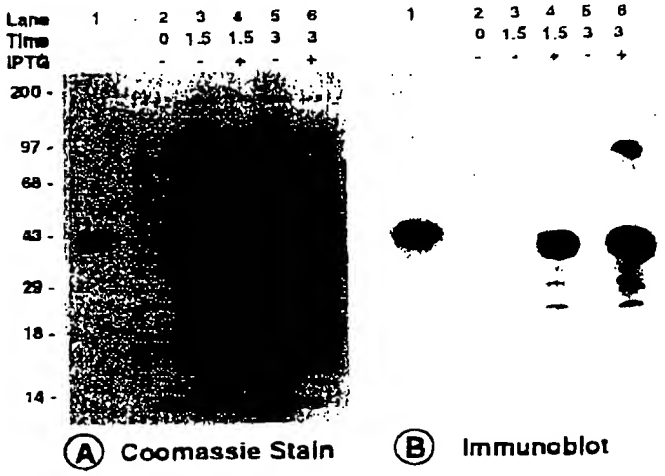


FIG. 2. Induction of SPARC production is dependent on inactivation of the *lac* repressor. Recombinant BL21(DE3) *E. coli* that contained pSPARCwt were grown in liquid culture until the $OD_{600}$ was 0.6–0.8. IPTG was added to 1 mM at Time 0. One-milliliter aliquots of culture were removed at the times indicated, centrifuged at 16,000*g* for 1 min, and dissolved in 0.2 ml of SDS–PAGE sample buffer that contained 50 mM DTT. Equal volumes of each sample (0.02 ml) were fractionated by SDS–PAGE on 10% gels that contained 0.1% SDS. (A) SDS gel stained with Coomassie blue. (B) Autoradiogram from an immunoblot. Antibodies were specific for the C-terminal EF-hand region of rSPARC (16). SPARC:IgG complexes were detected with $^{125}$I-protein A. Lane 1, native SPARC from murine PYS cells (1 μg); lanes 2–6, total *E. coli* protein at times indicated; lane 2, 0 h; lane 3, 1.5 h without IPTG; lane 4, 1.5 h with IPTG; lane 5, 3 h without IPTG; lane 6, 3 h with IPTG.

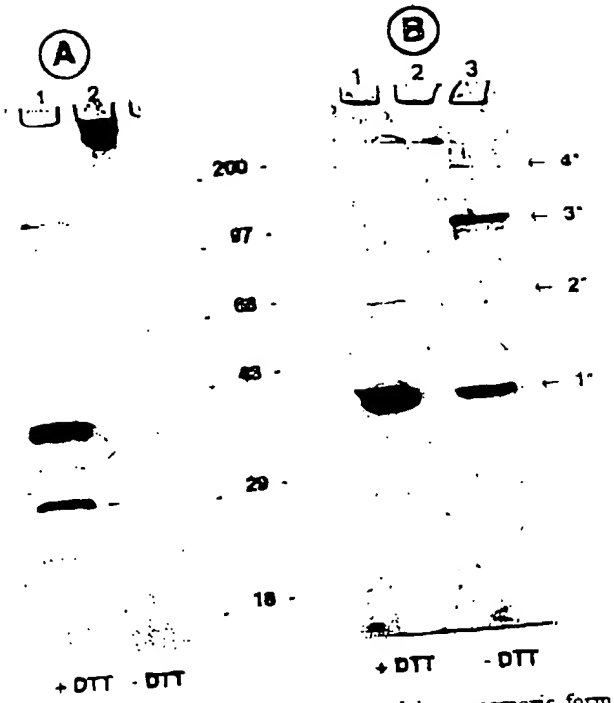


FIG. 3. Soluble rSPARC is synthesized in monomeric form. The recombinant protein was isolated by sequential chromatography on ion-exchange and nickel-chelate resins, respectively, as described under Experimental Procedures. Shown are Coomassie blue stains of purified protein after elution from the Ni–NTA column. The same quantity of protein (10 μl = 5 μg) was mixed with an equal volume of 0.125 M Tris–HCl (pH 6.8), 10% glycerol, 4% SDS, and 0.01% bromphenol blue. DTT was added to one protein sample to a final concentration of 50 mM. Both samples were incubated at 95°C for 10 min, cooled to room temperature, and resolved by a 10–20% polyacrylamide electrophoretic gel that contained 0.1% SDS. 1°, 2°, 3°, and 4° refer to the mono-, di-, tri-, and tetrameric forms of the recombinant SPARC protein, respectively. (A) Aggregation of rSPARC during dialysis. Bacterial lysate was dialyzed against ion-exchange equilibration buffer for 16 h at 4°C prior to chromatography. Lane 1, +DTT; lane 2, –DTT. (B) Resolution of rSPARC aggregates on DEAE-Sepharose resin. Bacterial lysate was incubated with ion-exchange slurry by gentle stirring for 16 h at 4°C, prior to ion exchange and Ni-chelate chromatography. Lane 1, +DTT; lane 2, blank; lane 3, –DTT.

ing dialysis against ion-exchange chromatography buffer. Proteins that remained soluble after dialysis were subjected to ion-exchange and, subsequently, to nickel-chelate chromatography. Analysis of the final isolation product by SDS–PAGE demonstrated that rSPARC polypeptide chains were aggregated (Fig. 3A). The isolated protein failed to enter the electrophoretic gel in the absence of DTT (Fig. 3A, lane 2). The addition of reducing agent prior to electrophoresis converted the aggregate to bands of apparent molecular masses of 32 and 28 kDa, which presumably represented a monomeric polypeptide chain and its degradation product, respectively. Since the 38- and the 28-kDa species both reacted with anti-SPARC immunoglobulins (data not shown), we concluded that the 28-kDa moiety was a degradation product of the 38-kDa band. However, we were unable to obtain an $NH_2$ sequence of the 28-kDa band by Edman degradation due to high background.

To avoid the precipitation of rSPARC during the purification process, we mixed the soluble bacterial extract directly with the anion-exchange resin for 16 h. This procedure proved to be critical, in that final preparations of soluble rSPARC consisted primarily of monomers (Fig. 3B). We compared the electrophoretic mobilities of rSPARC treated with/without DTT prior to SDS–PAGE. The major electrophoretic form of rSPARC in the presence of 50 mM DTT exhibited an apparent molecular mass of 34–38 kDa (Fig. 3B, lane 1). The amount of this monomeric form was reduced by approximately 33% in the absence of disulfide bond reduction (Fig. 3B, lane 3). We conclude that purified rSPARC was principally monomeric, with successively lesser amounts of trimeric, dimeric, and tetrameric forms, and that the protein had resolved into these configurations through interactions with ion-exchange resin.

*Nickel-chelate affinity chromatography.* We incorporated six His residues into the COOH-terminus of rSPARC to facilitate isolation of the expressed protein. Under the conditions described under Experimental Procedures, rSPARC bound to the nickel-chelate resin at an affinity greater than that of the contaminating bacterial proteins (21). Initial experiments were performed with metal-chelate affinity resins in which iminodiacetate (IDA) was used as the chelating ligand. The IDA ligand, coupled to agarose, functioned to coordinate $Ni^{2+}$ ions to the His residues of rSPARC. We observed that $Ni^{2+}$ ions, having six coordination sites, were only weakly bound to the matrix and were washed out during elution of our rSPARC samples with imidazole. Since the IDA ligand has only three chelating sites and cannot bind metal ions containing six coordination sites with reasonable affinity, we used metal-chelate resins with the chelating ligand nitrilotriacetate (NTA). This Ni–NTA metal-chelate resin has four chelating sites, allows a stable interaction between $Ni^{2+}$ and column matrix, and leaves two metal coordination sites free to interact with functional metal residues (21). Initially, we used imidazole to elute rSPARC from Ni–NTA resins. However, the intrinsic absorbance of this chemical at 280 nm proved problematic in the detection of rSPARC, and ultrapure commercial preparations that produced less background absorbance were prohibitively expensive. We subsequently realized that a reduction in pH during elution would protonate the His residues of rSPARC and allow a more favorable elution of the protein. As described under Experimen-

TABLE I

Purification of Soluble rSPARC from *Escherichia coli*[a]

| Fractions | Total proteins (mg) | Percentage of 34- to 38-kDa protein[b] (% area) | 34- to 38-kDa protein (mg) | Fold purification | Yield (%) |
|---|---|---|---|---|---|
| Soluble lysate | 825 | 23[c] | 190[d] | 1 | 100 |
| DEAE-Sepharose[e] | 56 | 51[d] | 29[d] | 2.2[d] | 15.3[d] |
| Ni–NTA[f] | 3.5 | 100 | 3.5 | 2 | 12.1 |

[a] Based on one representative experiment.
[b] Calculated by integration of densitometric scans of photographic negatives.
[c] Estimate is high due to number of proteins in this molecular size range.
[d] Calculation depends on the percentage of area for the 34- to 38-kDa protein (footnote c).
[e] From a 0.2–0.3 M NaCl eluate.
[f] From a pH 5.3 eluate.



tal Procedures, rSPARC bound tightly to the Ni–NTA resin at pH 7.8 because the $pK_a$ of the imidazole moiety of His is 6.8. Under washing conditions at pH 6.0, contaminating proteins were efficiently removed at a flow rate of 0.5 ml $min^{-1}$ (data not shown). At pH 5.3, rSPARC was specifically eluted with less than 5% contaminating proteins. Figure 3B displays a representative electrophoretogram of a final preparation of rSPARC from a Ni–NTA column. We and others (22) have found that a low flow rate is essential for achievement of equilibrium during the washing or elution steps. Typical yields of soluble rSPARC that were eluted from this resin were 3–4 mg rSPARC per soluble extract derived from a fermentation of 1.3 liters.

We attempted to determine the yield of rSPARC at each step of the purification process. Table I shows the total milligrams and the milligrams of rSPARC at each step. We were unable to determine accurately the quantity of rSPARC in the starting soluble lysate material because of the high number of bacterial proteins present in this mixture. The percentage of area that we calculate for rSPARC by densitometric scanning of photographic negatives is likely to be higher than the true area because of numerous proteins that constitute the 34- to 38-kDa fraction on SDS–PAGE. However, at the DEAE-Sepharose step, rSPARC is the predominant protein of 34–38 kDa.

### *Isolation of Monomeric rSPARC by Gel Filtration Chromatography*

We occasionally noticed higher molecular weight forms of rSPARC in our final preparations, especially after several freeze–thaw cycles. To remove dimers, trimers, and oligomers from monomeric rSPARC, we fractionated the eluates from the Ni–NTA affinity resin on a Superdex-70 gel filtration resin. Figure 4A displays a representative chromatogram and demonstrates that the protein eluting between 55 and 60 ml is the monomeric form of rSPARC, as judged by nonreducing SDS–PAGE (Fig. 4A, fraction 37). A dimeric form of rSPARC elutes just after the void volume at fraction 32 and a degradation fragment of 22 kDa at fraction 47. Configurations of rSPARC greater than dimers do not enter the gel matrix and are present in the void volume (not shown).

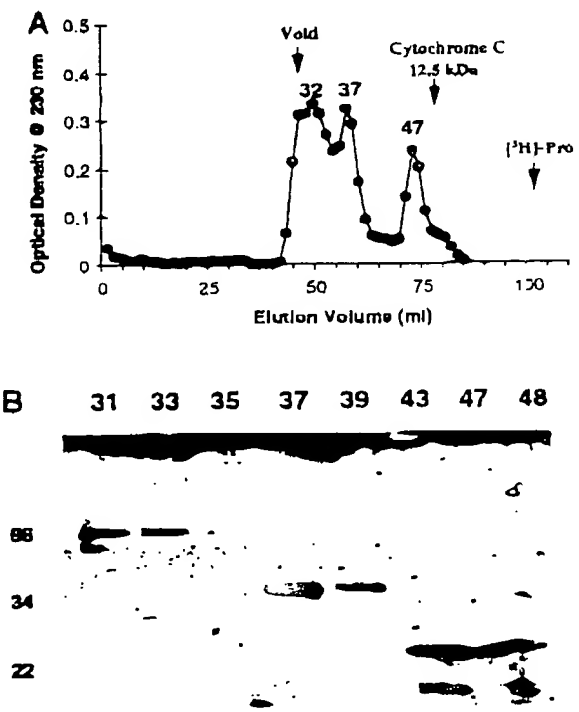


FIG. 4. Isolation of monomeric rSPARC by gel filtration. Protein eluates from Ni–NTA affinity chromatography were applied to a 1.6 × 60-cm column of Superdex-70 equilibrated in 50 mM Tris–HCl (pH 8.0) containing 0.15 M NaCl. (A) Fractions of 1.25 ml were monitored at $A_{280}$ prior to (B) analysis by SDS–PAGE. Protein samples were denatured with 2% SDS, subjected to electrophoresis under nonreducing conditions, and stained with Coomassie blue. The column fraction number is indicated at the top of each well, and $M_r$ values are given to the left.

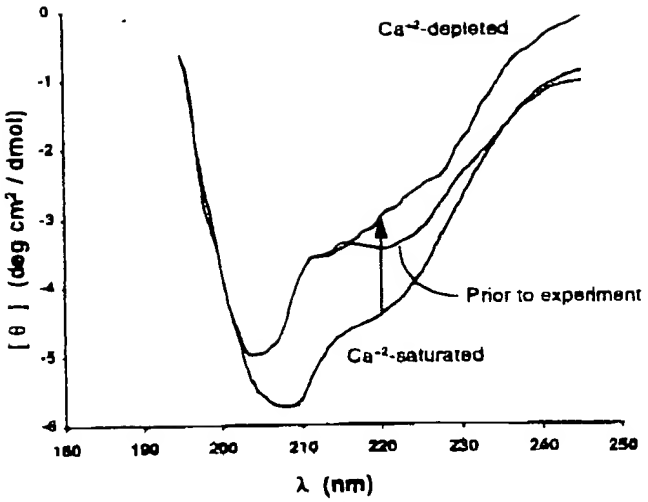


FIG. 5. Induction of $\alpha$-helix in rSPARC by $Ca^{2+}$. Fifty micrograms of rSPARC in a Chelex-treated buffer of 10 mM Tris–HCl (pH 7.8) and 0.15 M NaCl was placed in a cell of pathlength 0.1 mm, and the solution was scanned from 290 to 195 nm in a circular dichroism spectrometer. Shown is the average of six scans in the absence of exogenous calcium (prior to experiment). Successive samples, saturated with and depleted of $Ca^{2+}$, are indicated. Shown are spectra after the subtraction of the spectrum of the buffer solution. The arrow refers to the range of EDTA titration described in Fig. 6.

*Analysis of $NH_2$- and COOH-Termini*

We subjected a preparation of soluble rSPARC to amino acid sequencing by the Edman degradation procedure and acquired the following sequence: P-Q-Q-E-A-L-P-D-E-T-E-V-V-E-E. Since the sequence obtained was not blocked, we concluded that the first 2 $NH_2$-terminal amino acids, formyl-Met and Ala, had been removed by bacterial processing. The remaining 15 residues match the predicted, processed $NH_2$-terminal end of cDNA sequences encoding human SPARC (13, 14, 23). Together with our observation that the protein bound strongly to the Ni–NTA nickel-chelate resin through its COOH-terminal $(His)_6$ motif, we have concluded that preparations of rSPARC are essentially intact. Based on 293 residues present in the isolated protein, the $M_r$ was calculated as 33,692 and the isoelectric point as 4.71.

*Conformational Activities of rSPARC Are Induced by $Ca^{2+}$*

*Circular dichroism spectroscopy (CD).* We monitored secondary structure and conformational changes in rSPARC as a function of $Ca^{2+}$ concentration by CD (Fig. 5). The minimum observ d in the CD spectrum of a polypeptide at 220 nm is indicativ of $\alpha$-helical content. The addition of $Ca^{2+}$ ion to a rSPARC sample resulted in the increase of the mean residue ellipticity ($\theta$) at 220 nm from $-3.4550$ to $-4.419$ deg · cm$^2$ · dmol$^{-1}$. This increase of 28% indicated that a significant amount of $\alpha$-helicity was obtained. rSPARC therefore bound $Ca^{2+}$, and this interaction resulted in a change in $\alpha$-helical content that could be monitored by CD. This result further demonstrates that the bacterial cytoplasm does not contain a sufficient pool of free $Ca^{2+}$ to allow full occupation of the high- and low-affinity $Ca^{2+}$ binding sites of rSPARC, but that the isolated protein had folded into a conformation from which it could acquire further $\alpha$-helicity. This increase in ellipticity could be reversed by EDTA, a previously reported feature in the CD studies of native SPARC (24, 25, 34).

We studied the transition in conformation when bound $Ca^{2+}$ was gradually removed from rSPARC and native murine SPARC by titration with EDTA (Fig. 6). The data show that both proteins display a titration curve indicative of cooperativity. Although the concentrations of both the recombinant and the murine protein samples were equivalent, the lesser magnitude of the murine SPARC signal was likely due to the presence of stabilizing serum albumin in these preparations. These transitional changes in SPARC are in agreement with titration curves previously reported for native SPARC isolated from murine parietal endoderm cells (24) or from the Engelbreth–Holm–Swarm murine tumor (25, 34). rSPARC thus displayed a conformational activity highly similar to that of the native protein. A comparison of $\Delta[\theta]_{220}$ between recombinant and murine forms of SPARC is listed in Table II (see below for further discussion).

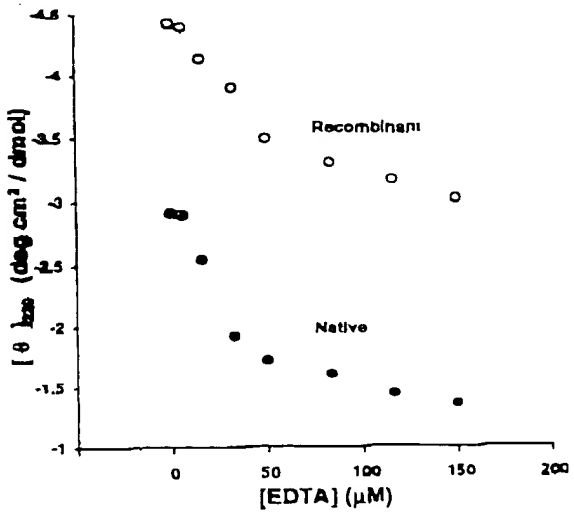


FIG. 6. $Ca^{2+}$-dependent conformational transition of rSPARC and native SPARC. The change of circular dichroism ellipticity at 220 nm (see arrow in Fig. 5) was monitored as a function of increasing amounts of EDTA added to rSPARC (○) or native, murine SPARC (●).

TABLE II

Comparison of the $Ca^{2+}$-Induced Changes in Human SPARC and Murine SPARC

| Protein (species) source of isolation | Experimental method[a] | $\Delta F_{336}$ (%) | $\Delta[\theta]_{220}$ (%) | $K_d$ ($\mu$M) |
|---|---|---|---|---|
| rSPARC (human) bacteria | EDTA titration | 44% (Fig. 8) | 32[b] (Fig. 6) | 100 (Fig. 8) |
| SPARC (murine) parietal endoderm cells | EDTA titration | 36% (Fig. 8) | 54[b] (Fig. 6) | 45 (Fig. 8) |
| | | | 35[c] (Ref. 4) | 50 (Ref. 24) |
| SPARC (murine) Engelbreth–Holm–Swarm tumor | $Ca^{2+}$ titration | 34% (Ref. 34) | 34[d] (Ref. 34) | 0.6 (Ref. 34) |
| rSPARC human kidney cells | $Ca^{2+}$ titration | 95% (Ref. 10) | 36[e] (Ref. 10) | 0.08 (Ref. 10) |

*Note.* $\Delta F$ and $\Delta[\theta]$ were calculated from signals of $Ca^{2+}$-free and $Ca^{2+}$-saturated proteins according to the formula (free − saturated)/saturated (10). $K_d$ were calculated from midpoint inflections of fluorescence signals (Fig. 8).

[a] A fundamental difference exists in how SPARC was prepared prior to spectroscopic determination between this report and the data from Ref. 10 and 34. For EDTA titrations, SPARC was saturated with $Ca^{2+}$ and unbound $Ca^{2+}$ was removed. For $Ca^{2+}$ titrations, SPARC was passaged through an EDTA chromatography column.

[b] EDTA present at 100-fold excess over protein, 150:1.5 mM (Fig. 6).

[c] EDTA present at 91-fold excess over protein, 1000:11 $\mu$M (24).

[d] EDTA present at 2-fold molar excess over protein (34).

[e] EDTA present at 2-fold molar excess over protein (10).



*Intrinsic fluorescence emission spectroscopy.* We analyzed the interaction of $Ca^{2+}$ with rSPARC by monitoring of the fluorescence emission profile of intrinsic Trp residues (Figs. 7 and 8). The comparison of the emission profiles between the aromatic amino acids and rSPARC at varying ultraviolet wavelengths indicated that Trp was the primary contributor to the observed signal of the protein (not shown). Figure 7 demonstrates that 280 nm is the optimal $\lambda_{ex}$ for the recombinant protein because (a) the greatest amount of fluorescence intensity was observed and (b) the emission spectrum was resolved optimally from that contributed by the intrinsic Rayleigh scattering of the sample and cuvette. The intensity of the emission profiles of rSPARC was therefore dependent on an aromatic amino acid, primarily Trp.

When a sample of rSPARC, identical to that shown in Fig. 7A, was saturated with $Ca^{2+}$, and unbound cation was subsequently removed by gel filtration chromatography, the fluorescence intensity of the emission profile at each $\lambda_{ex}$ was found to decrease significantly (Fig. 7B). This decrease was consistent with fluorescent emission profiles for native SPARC (24), for native SPARC/BM-40 (25, 34), and for recombinant SPARC/BM-40 (10) in the presence and absence of $Ca^{2+}$.

The gradual removal of bound $Ca^{2+}$ from rSPARC and murine SPARC was achieved by titration with EDTA. Figure 8 demonstrates that over a 6000-fold

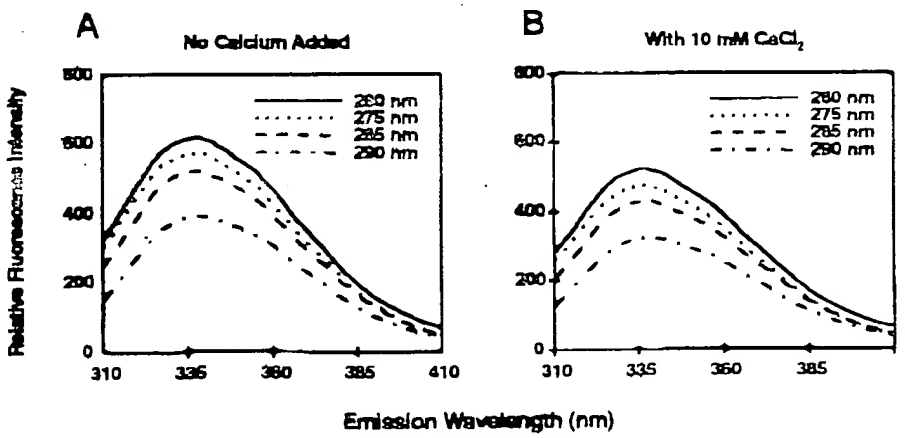


FIG. 7. Fluorescence emission spectra of rSPARC are reduced by $Ca^{2+}$. The optimal excitation wavelength of a purified, soluble preparation of rSPARC was determined for fluorescence emission spectroscopy. Resolution slitwidths were 6 nm for excitation and 7 nm for emission. Scan speed was 200 nm/min. Shown are spectra after subtraction of spectra corresponding to buffer. The optimal wavelength for excitation of the aromatic amino acids (primarily Trp) of rSPARC was 280 nm. (A) Spectra of rSPARC (30 $\mu$g) isolated from bacteria. (B) Spectra of rSPARC (30 $\mu$g) after addition of 10 mM $CaCl_2$.

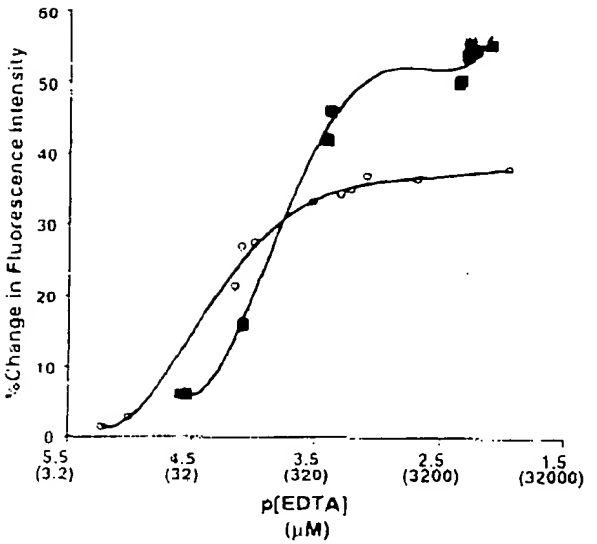


FIG. 8. Removal of bound $Ca^{2+}$ alters the intrinsic fluorescence of rSPARC and native SPARC. Proteins were dissolved into a 0.05 M Tris · HCl buffer (pH 8.0) containing 0.15 M NaCl and 0.05 M $CaCl_2$. $Ca^{2+}$ not bound to the protein was removed by size-exclusion chromatography. Proteins were diluted to 1.5 $\mu$M in the same buffer that contained varying concentrations of EDTA. Shown is a plot of the percentage of change in intrinsic fluorescence at 336 nm as a function of pEDTA. The data for rSPARC (■) were plotted as a fourth-order polynomial function according to $y = 11.642x^4 + 182.48x^3 + 721.41x^2 + 1476.4x + 1161.6$; $R^2 = 0.9939$. The data for murine SPARC, isolated from mouse parietal endoderm cells, were plotted (○) as a sixth-order polynomial function according to $y = 0.1225x^6 + 1.5178x^5 + 7.1159x^4 + 5.3065x^3 - 65.766x^2 + 132.52x - 47.945$; $R^2 = 0.9938$.

concentration range of EDTA ($1.1 \times 10^{-8}$–$1.8 \times 10^{-6}$ M), the peak height of the emission spectra of rSPARC at 336 nm (closed squares) was diminished by the chelating agent in a manner consistent with that observed for murine SPARC (open circles). From these data, we estimated the dissociation constants for rSPARC and murine SPARC as 100 and 45 $\mu$M, respectively. These values are in agreement with previously reported values for the protein studied in analogous assays and indicate that rSPARC exists in a conformation that is highly similar to that of the native protein (Table II). $Ca^{2+}$ therefore influenced the fluorescence emission spectrum of the intrinsic Trp residues in rSPARC, through either an alteration of the environment around this amino acid or the quenching of Trp fluorescence by neighboring amino acids.

The methodology used for purification of SPARC prior to spectroscopic determination most likely accounts for the differences in the values noted in Table II. For EDTA titrations, SPARC was saturated with $Ca^{2+}$ and unbound $Ca^{2+}$ was removed (values in the first two rows of Table II). For $Ca^{2+}$ titrations, SPARC was passed through an EDTA chromatography column (10, 31) (values in rows 3 and 4 of Table II). Because of this difference in experimental protocol, and because of inherent differences in expression systems (see Discussion), we cannot compare directly the values obtained for $\Delta F_{max}$ and $\Delta[\theta]_{max}$ of human SPARC expressed in bacteria (this report) with human SPARC expressed in kidney cells (10).

### *rSPARC Displays Biological Activities*

*Inhibition of endothelial cell spreading by rSPARC.* We tested our isolated soluble rSPARC in an assay of cell spreading. As shown in Fig. 9A, purified rSPARC inhibited the spreading of BAE cells in a concentration-dependent manner. Cell rounding was observed at 40 $\mu$g/ml (~2 $\mu$M) added protein. This result is consistent with the well-known property of native SPARC to inhibit the spreading of newly plated BAE cells, fibroblasts, and smooth muscle cells (9).

*Specific binding to endothelial cells.* Radiolabeled rSPARC bound specifically to confluent monolayers of BAE cells at pH 7.4 in a concentration-dependent manner (Fig. 9B, black bars). Competition was achieved with unlabeled SPARC peptide 4.2, and specific binding was calculated by subtraction of nonspecific binding from total binding (19). In separate experiments, we observed that a 200-fold excess of unlabeled peptide 4.2 effectively competed for the binding of $^{125}$I-peptide 4.2 to BAE cells (not shown). For comparison, a previously reported experiment of the binding of $^{125}$I-murine SPARC to BAE cells at pH 7.4 is shown in Fig. 9C. These results clearly indicate that both rSPARC and native SPARC are able to bind specifically to endothelial cells (19).

*Presence of bacterial endotoxin.* We measured the levels of contaminating bacterial endotoxin in six different preparations of rSPARC (Table III and Fig. 10). After correcting for the contribution of endotoxin by media and buffers, we calculated that preparations of rSPARC contained 0.014–0.018 ng endotoxin/$\mu$g rSPARC. This ratio is below the threshold *in vitro* that elicits a significant change in the gene expression of BAE cells (26) and causes cellular detachment (e.g., 27). Our protocol for anti-adhesion assays is not directly comparable to studies by Harlan *et al.* (27), which showed that a 20-min exposure to 1 $\mu$g/ml endotoxin was sufficient to induce 81% detachment of BAE monolayers. The maximum amount of endotoxin (0.72 ng) present in 40 $\mu$g of rSPARC was not observed to detach BAE cells during the rounding assay shown in Fig. 7A.

## DISCUSSION

The isolation of SPARC/osteonectin/BM-40 from tissues and conditioned media has resulted in preparations that are often contaminated by growth factors and/or proteinases, with attendant losses of concentra-

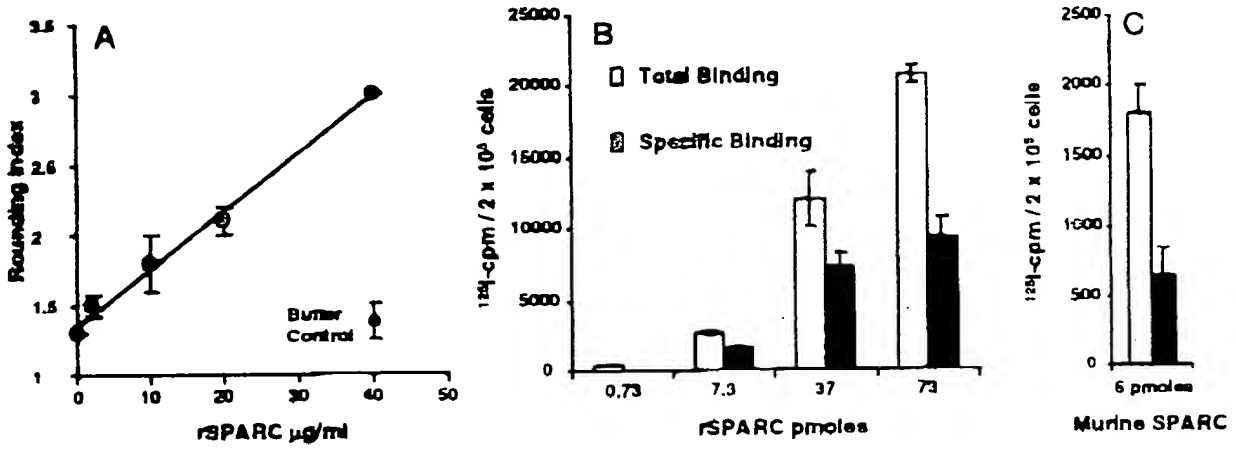


FIG. 9. Biological activity of rSPARC. (A) Anti-adhesion. rSPARC (100 μg/ml) was dialyzed against DMEM that contained 1% FBS and was added to tissue culture plastic wells at final concentrations of 0, 2, 10, 20, and 40 μg/ml (0, 0.06, 0.30, 0.6, and 1.2 μM). BAE cells ($10^4$) were subsequently added to each well. After 2 h, the cells in each group were evaluated according to a rounding index (16). A score of 1 represents a culture with only spread cells, whereas a culture with increasing numbers of unspread and round cells would approach the maximum rounding index of 3. Values were calculated for two independent cultures and graphed as means +/− standard error. A control sample contained the same volume fraction of Ni–NTA elution buffer (pH 5.3). (B) Binding studies. Iodinated rSPARC was incubated at pH 7.4 with confluent BAE cells grown in 24-well plates as described under Experimental Procedures. Bound $^{125}$I-rSPARC was competed by a 200-fold excess of SPARC peptide 4.2 (representing amino acids 255–273) and specific binding was obtained by subtraction of nonspecific binding from total binding (17). Bars represent the means +/− standard deviation of 3 wells. Binding is expressed as the mean $^{125}$I cpm/well containing $2 \times 10^5$ cells ± SD. The specific activity of $^{125}$I-rSPARC was 621 cpm/fmol. (C) Binding of murine SPARC to BAE cells. Data taken from Ref. 19. BAE cells were grown in 24-well plates until confluent. To each well, 6 pmol of $^{125}$I-murine SPARC was incubated in binding buffer at pH 7.4 ± a 50-fold molar excess of unlabeled SPARC for 2 h at 4°C. Cultures were washed, solubilized in 1 N NaOH, and counted. Binding is expressed as the mean cpm bound/$2.5 \times 10^5$ cells ± SE. The specific activity of murine SPARC was 1085 cpm/fmol. Bars are labeled as in B.

tional and biological activities. Whereas highly purified preparations of SPARC have proven valuable for delineation of protein activity, quantities of the protein sufficient for extensive biophysical and structural analyses have not been available. We have attempted to address these problems by expression of SPARC in bacteria and we report here that isolations of the purified protein are biologically active and correctly folded.

TABLE III

Endotoxin Levels in rSPARC Preparations

| Sample | ng endotoxin/μg rSPARC[a] |
|---|---|
| rSPARC-1[b] | 0.018[c] |
| rSPARC-2[b] | 0.018[c] |
| rSPARC-3[d] | 0.016[e] |
| rSPARC-4[b] | 0.014[c] |

[a] Endotoxin level determined from standard curve (Fig. 10). Concentration of rSPARC was determined by BCA assay.
[b] rSPARC in DMEM solution.
[c] Value obtained after subtraction of endotoxin levels in DMEM + Ni–NTA elution buffer.
[d] rSPARC in RPMI solution.
[e] Value obtained after subtraction of endotoxin levels in RPMI + Ni–NTA elution buffer.

Our finding that the resolution of rSPARC aggregates on ion-exchange resins could increase our yields from <0.2 μg to >3 mg per 1.3 liters of fermentation culture is consistent with the results of Hoess *et al.* (32), who showed that treatment of bacterial lysates with ion-exchange resin could solubilize three different recombinant proteins into biologically active forms. Since a significant amount of rSPARC (~50%) is found in the insoluble fraction of the lysate, we are currently assessing whether this technique will resolubilize and correctly fold the insoluble form of the protein. Because the synthesis of rSPARC RNA is under the control of the strong T7 bacteriophage promoter, we have observed rapid accumulation of recombinant protein within 15 min of the addition of IPTG. We believe that our system becomes saturated rapidly with rSPARC protein; hence, there is a sequestration of protein into the insoluble fraction. The soluble extract that we prepare must therefore contain a significant fraction of protein that is marginally soluble due to varying degrees of tertiary structure. Through the interaction of rSPARC with DEAE-Sepharose resin, we surmise that the protein adopts a thermodynamically stable arrangement of disulfide bonds and hydrophobic interactions that leads to the formation of soluble monomers over the 16–24 h that rSPARC is in contact with the

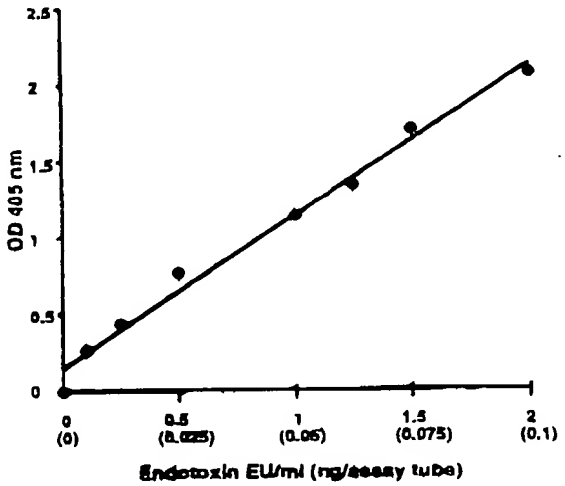


FIG. 10. Determination of bacterial endotoxin concentration. Known quantities of endotoxin were diluted over the indicated range, incubated with *Limulus* amoebocyte lysate (15) for 10 min at 37°C, and measured photometrically by the enzymatic cleavage of *p*-nitroaniline (pNA) from Ac-Ile-Glu-Ala-Arg-pNA at 405 nm.

resin. However, rSPARC does not necessarily require DEAE-Sepharose resin for the resolution of aggregates, because insoluble, denatured rSPARC can be refolded into soluble monomers by gradient dialysis and by cystine isomerization (J. Bassuk *et al.*, manuscript in preparation). Although soluble rSPARC monomers exist in a conformation highly similar to that of the native protein isolated from mammalian cell culture, confirmation of this point awaits the determination of the 3-dimensional structure of SPARC.

Eukaryotic proteins expressed in bacteria are not glycosylated. Since SPARC/BM-40/osteonectin is known to contain complex oligosaccharide chains that vary in length and composition (11), it was possible that the lack of carbohydrate in rSPARC would diminish its biological activity. Kelm and Mann (11) reported that changes in the oligosaccharide structure of SPARC altered the binding to collagen. In contrast, Pottgieser *et al.* (10) reported that deglycosylation of SPARC/BM-40 had no effect on the binding to collagen IV. A mutant rSPARC lacking the known N-linked carbohydrate attachment site was still active in cellular anti-adhesion assays (38). Our results are consistent with the latter two reports. Clearly, a functional activity for the oligosaccharide side chain(s) of the native protein has yet to be identified.

SPARC/BM-40 exhibits a significant change in $\alpha$-helical content upon removal of bound $Ca^{2+}$ ion (10, 24, 25). We used CD spectroscopy to analyze an isolated preparation of rSPARC and observed a change in the mean residue ellipticity ($\Delta\theta_{220}$) of 32% upon the addition of $Ca^{2+}$. This increase in $\Delta\theta_{220}$ was consistent with the acquisition of $\alpha$-helicity, and we concluded that rSPARC responds to the addition of exogenous $Ca^{2+}$ in a manner similar to that of native SPARC. We are unable to add $Ca^{2+}$ to rSPARC during the initial stages of purification because $CaPO_4$ is insoluble in the phosphate buffer system used. Newly synthesized rSPARC might therefore be depleted of $Ca^{2+}$ at the low-affinity, high-capacity binding sites present in the Glu-rich, $NH_2$-terminal 50 amino acids, a consequence, perhaps, of limiting amounts of free $Ca^{2+}$ in the bacterial cytoplasm.

We also measured the fluorescence emission profile of rSPARC in response to excitation with ultraviolet light. After correction for the intrinsic Raman spectra of water present in all buffers and solutions, the maximal fluorescence emission intensity in the presence of 10 mM $CaCl_2$ was 85% of the signal that corresponded to the protein with no added $Ca^{2+}$. A value of 95% was reported for human recombinant SPARC/BM40 (10). We conclude that bacterial rSPARC, like its mammalian counterpart, alters the environment of its intrinsic Trp residues in response to $Ca^{2+}$. Within the EF-hand of human SPARC is a single Trp (residue 269) that is likely to account, at least in part, for the fluorescence emission signal. We also estimated dissociation constants for rSPARC and murine SPARC from the EDTA-titration curves by location of the $\frac{1}{2}$ maximal change on the ordinate of Fig. 8 (% Change in Fluorescence) and calculation of the corresponding position on the abscissa (pEDTA). We found, with reasonable error attributable to methodogy, that the values of 100 and 45 $\mu$M, respectively, agreed with the literature values obtained by EDTA titration. When we compared our experimental values with the data from Mayer *et al.* (34) and Pottgiesser *et al.* (10), we detected a fundamental difference in methodology in fluorescence titrations of SPARC. In these references, the authors first remove all $Ca^{2+}$ by passage of SPARC/BM-40 through an EDTA-resin and subsequently titrated the $Ca^{2+}$ back. Under these conditions, those authors obtained a value of 0.6 $\mu$M for murine SPARC isolated from the Engelbreth–Holm–Swarm tumor. We conclude that our EDTA titrations are not directly comparable to $Ca^{2+}$ titrations, but that the overall response of the protein to $Ca^{2+}$ is fundamentally the same. Differences in binding activities may not be due only to conformational changes, but to posttranslational changes (such as glycosylation or phosphorylation). Surface charge and long-range interactions (10) might also have profound effects on $Ca^{2+}$ affinity.

Since a bacterial expression system should generate biologically active material, we evaluated whether rSPARC could alter the morphology and adhesive properties of aortic endothelial cells. Our results demonstrate that the protein inhibited cell spreading on tissue culture plastic in a concentration-dependent man-

ner. This observation is in agreement with the known property of native SPARC as an anti-spreading factor (9). Our determination that rSPARC preparations contained <0.018 ng endotoxin/μg rSPARC allowed us to conclude that endotoxin was not responsible for the change in morphology and adhesion of BAE cells, because this level of endotoxin is below the threshold reported to elicit alterations in gene expression (26), attachment (27), and viability (27) of these cells. We also found that rSPARC specifically bound to confluent monolayers of endothelial cells, a result consistent with activities reported for native SPARC purified from murine cells (19). The binding of rSPARC was competed by the 20-residue SPARC peptide 4.2, a sequence that comprises the EF-hand loop present at the C-terminus of the protein. We have also found that rSPARC reduced the incorporation of [$^3$H]thymidine by rat glomerular mesangial cells and human smooth muscle cells *in vitro* (J. Bassuk *et al.*, manuscript in preparation; S. Funk *et al.*, manuscript in preparation).

We are currently improving the yields and folding of biologically active rSPARC through the expression and transport of the protein into the bacterial periplasmic space, a compartment that is less reducing for disulfide bonds than the bacterial cytoplasm. The availability of recombinant SPARC that is biologically active and correctly folded now allows us to discern the structure/function relationships of this multifunctional extracellular protein.

## ACKNOWLEDGMENTS

We thank Pieter Oort (Zymogenetics, Inc., Seattle, WA) for the PCR amplification of human SPARC, Dr. Cathy-Lofton Day (Zymogenetics) for the human endothelial cell cDNA library, Michael Sun and Eleonor Schneider for provision of bacterial fermentation pastes, Laurie Braun for technical assistance, and Dr. Trisha Davis for recombinant calmodulin.

## REFERENCES

1. Lane, T. F., and Sage, E. H. (1994) *FASEB J.* 8, 163–173.
2. Timpl, R., and Aumailley, M. (1993) *in* Molecular and Cellular Aspects of Basement Membranes (Rohrbach, D. H., and Timpl, R., Eds.), pp. 211–236, Academic Press, Orlando.
3. McVey, J. H., Nomura, S., Kelly, P., Mason, I. J., and Hogan, B. L. M. (1988) *J. Biol. Chem.* 263, 11111–11116.
4. Mason, I. J., Taylor, A., Williams, J. G., Sage, H., and Hogan, B. L. M. (1986) *EMBO J.* 5, 1465–1472.
5. Mann, K., Deutzmann, R., Paulsson, M., and Timpl, R. (1987) *FEBS Lett.* 218, 167–172.
6. Bolander, M. E., Young, M. F., Fisher, L. W., Yamada, Y., and Termine, J. D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2919–2923.
7. Lane, T. F., Iruela-Arispe, M. L., Johnson, R. S., and Sage E. H. (1994) *J. Cell Biol.* 125, 929–943.
8. Nischt, R., Pottgiesser, J., Krieg, T., Mayer, U., Aumailley, M. and Timpl, R. (1991) *Eur. J. Biochem.* 200, 529–536.
9. Sage, H., Vernon, R. B., Funk, S. E., Everitt, E. A., and Angello, J. (1989) *J. Cell. Biol.* 109, 341–356.
10. Pottgiesser, J., Maurer, P., Mayer, U., Nischt, R., Mann, K., Timpl, R., Krieg, T., and Engel, J. (1994) *J. Mol. Biol.* 238, 563–574.
11. Kelm, R. J., and Mann, K. G. (1991) *J. Biol. Chem.* 266, 9632–9639.
12. Raines, E. W., Lane, T. F., Iruela-Arispe, M. L., Ross, R., and Sage, E. H. (1992) *Proc. Natl. Acad. Sci. USA* 89, 1281–1285.
13. Lankat-Buttgereit, B., Mann, K., Deutzmann, R., Timpl, R., and Krieg, T. (1988) *FEBS Lett.* 236, 352–356.
14. Villarreal, X. C., Mann, K. G., and Long, G. L. (1989) *Biochemistry* 28, 6483–6491.
15. Young, N. S., Levin, J., and Prendergast, R. A. (1972) *J. Clin. Invest.* 51, 1790–1796.
16. Sage, E. H., Basauk, J. A., Yost, J. C., Folkman, M. J., and Lane, T. F. (1995) *J. Cell. Biochem.* 57, 127–140.
17. Lane, T. F., and Sage, E. H. (1990) *J. Cell Biol.* 111, 3065–3076.
18. Bowen-Pope, D. F., and Ross, R. (1985) *Methods Enzymol.* 109, 69–100.
19. Yost, J. C., and Sage, E. H. (1993) *J. Biol. Chem.* 268, 25790–25796.
20. Hughes, R. C., Taylor, A., Sage, H., and Hogan, B. L. M. (1987) *Eur. J. Biochem.* 163, 57–65.
21. Hochuli, E., Dobeli, H., and Schacher, A. (1987) *J. Chromatogr.* 411, 177–184.
22. Schmitt, J., Hess, H., and Stunnenberg, H. G. (1993) *Mol. Biol. Rep.* 18, 223–230.
23. Swaroop, A., Hogan, B. L. M., and Francke, U. (1988) *Genomics* 2, 37–47.
24. Engel, J., Taylor, W., Paulsson, M., Sage, H., and Hogan, B. (1987) *Biochemistry* 26, 6958–6965.
25. Maurer, P., Mayer, U., Bruch, M., Jeno, P., Mann K., Landwehr, R., Engel, J., and Timpl, R. (1992) *Eur. J. Biochem.* 205, 233–240.
26. Hasselaar, P., Loskutoff, D. J., Sawdey, M., and Sage, E. H. (1991) *J. Biol. Chem.* 266, 13178–13184.
27. Harlan, J. M., Harker, L. A., Reidy, M. A., Gajdusek, C. M., Schwartz, S. M., and Striker, G. E. (1983) *Lab. Invest.* 48, 269–274.
28. Adams, J. M. (1968) *J. Mol. Biol.* 33, 571–589.
29. Livingston, D. M., and Leder, P. (1969) *Biochemistry* 8, 435–443.
30. Vogt, V. M. (1970) *J. Biol. Chem.* 245, 4760–4769.
31. Ben-Bassat, A., Bauer, K., Chang, S. Y., Myambo K., Boosman, A., and Chang, S. (1987) *J. Bacteriol.* 169, 751–757.
32. Hoess, A., Arthur, A. K., Wanner, G., and Fanning, E. (1988) *Bio/Technology* 6, 1214–1217.
33. Yost, J. C., Bell, A., Seale, R., and E. H. Sage (1994) *Arch. Biochem. Biophys.* 314, 50–63.
34. Mayer, U., Aumailley, M., Mann, K., Timpl, R., and J. Engel. (1991) *Eur. J. Biochem.* 198, 141–150.